BAYES SEQUENTIAL DESIGN PROCEDURES:

THEORY AND APPLICATIONS

A thesis

submitted in fulfilment

of the requirements for the Degree

of

Doctor of Philosophy in Mathematics

in the

University of Canterbury

by

H.P. Edwards

University Of Canterbury,

Christchurch, New Zealand

November 1978

To my wife Joy

for her patience, understanding

and encouragement throughout the

preparation of this thesis

CONTENTS

CHAPTER PAGE

ABSTRACT

This thesis considers the problem of the sequential design of experiments from a Bayesian standpoint. Viewed as a sequential decision problem with a choice of experiments at each stage, the Bayes (i.e. optimal) sequential design procedure is constructively determined for the case when the number of experimentation stages is bounded. Under reasonably general conditions, it is shown that the Bayes sequential design procedure exists in the unbounded case and may be taken as the appropriate limit of optimal procedures in the bounded case. Also, an essentially complete class of sequential design procedures is characterised using generalised definitions of statistical sufficiency.

The constructive theory of Bayes sequential design procedures is then applied to two multiple decision problems, an identification problem and a ranking problem. Whilst the former illustrates the power of using the optimal procedure, the latter reveals the formidable computational difficulties involved in carrying out the backward induction for a non-trivial problem. Consequently, an alternative sequential design procedure is proposed, which, although based on the optimal procedure, is nevertheless easy to compute in most situations of interest. By modifying the goal of the experimenter in an appropriate fashion, it is shown that the alternative procedure possesses an optimality property for an identification problem concerning the parameters of two binomial populations. Finally, the results of computer simulation of the alternative procedure's performance in this problem and a similar problem concerning the parameters of two Normal populations are presented, which suggest that the procedure is superior to those procedures currently proposed for the problem. The computer programs used are reproduced in an appendix.

CHAPTER I

INTRODUCTION

1. THE SEQUENTIAL DESIGN OF EXPERIMENTS : ITS ORIGINS

The theory of sequential design of experiments is a comparatively recent branch of statistics, being the outgrowth of two seemingly unrelated fields, experimental design and sequential decision theory. The first of these, experimental design, arose from the need in agriculture and the biological sciences to analyse proposed schemes of experimentation in order to ensure that any statistical inference made on the outcome of such experimentation was, in fact, valid. For example, if it is desired to compare several varieties of wheat by planting one plot for each variety and observing the resultant yields, then the reasonable inference that the best variety is that one associated with the highest yield is not valid. It is possible that other factors such as soil fertility and moisture content could result in an inferior variety giving rise to a larger yield than that of the best variety. However, by designing the experiment so that all such factors are the same for each plot, or by siting the plots so that the effects of these factors cancel one another out, it is possible to make valid conclusions about the true yield of the varieties, such as rankings or estimates of these values etc. The theory and methodology of experimental design was first formulated by R.A. Fisher in the 1930's.

Sequential decision theory, on the other hand, is a more recent branch of statistics, and it is from the rapid growth in research in this area that the mathematical problem of the sequential design of experiments arose. From the experimenter's point of view, a statistical sequential decision problem may be described as follows: the experimenter is interested in some phenomenon or state of nature about which he is uncertain.

In order to increase his knowledge, he can observe the random outcome of an experiment (henceforth referred to as "taking an observation") which may be regarded as imparting information about the unknown phenomenon. Furthermore, he is able to continue taking observations until, at some stage, he is satisfied that he has acquired sufficient information about the phenomenon for his purposes, at which point he stops taking observations. The sequential decision problem for the experimenter is to find a rule which, after each observation, tells him whether he has sufficient information, in which case he stops, or whether more information is required, in which case he takes at least one more observation; when stopping occurs, it must also tell him which decision to make concerning the phenomenon. The problem is truly sequential because the number of observations required is never known in advance, being determined by the observed outcomes as the experimentation process continues.

Although the sequential decision problem as described above would seem to be a reasonable and useful problem to consider, little work appeared on the problem until the late 1940's when Wald published his celebrated work on sequential analysis (Wald (1947)) and statistical decision theory (Wald (1950)). His results opened up a vast new area of research in mathematical statistics, and consequently the problem of sequential design of experiments was soon considered. This is a generalisation of the sequential decision problem in which, at each stage of the experimentation process, the experimenter has available a choice of different experiments to perform. Thus, if he decides to take another observation, he must then decide which experiment to perform. Since some experiments are potentially more informative than others, a "good" choice of experiment can result in a large gain of information, and consequently a saving in the number of observations required. The experimental design process is also sequential, as the choice of the "best" experiment at each stage will depend on the outcome of the previous observations and experiment selections.

2. LARGE SAMPLE THEORY AND ASYMPTOTIC CONSIDERATIONS

The first study of the sequential design of experiments as a generalisation of the sequential decision problem was made by Chernoff (1959), who considered the usual Wald Sequential Probability Ratio Test (SPRT) of two simple hypotheses as a sequential decision problem, i.e. when no choice of experiment is available. Writing the hypotheses as

$$H_1 : f = f_1 \qquad \text{vs} \qquad H_2 : f = f_2 \tag{1.1}$$

where f is the density function of the independent and identically distributed (i.i.d.) observations, and f_1 and f_2 are two different known density functions, he showed that if the cost per observation, c, approaches zero, so that the expected sample size becomes large, then both the risk and the sample size of the procedure depend almost wholly on c and the Kullback-Leibler Information Numbers defined by

$$I_j = \int \log[f_j(x)/f_{3-j}(x)]\, f_j(x)\, dx, \quad j = 1,2, \tag{1.2}$$

and are relatively insensitive to the prior probabilities of the hypotheses and the costs of making an incorrect decision at termination. If the true hypothesis were known, then in the design situation where two experiments e_1 and e_2 are available, it would be possible, by comparing information numbers, to choose the more informative experiment at each stage. Of course the true hypothesis or state of nature is not known. However, providing the sample size is large, the maximum-likelihood estimate $\hat{\theta}_n$ is known to be close to the true state of nature θ, because of its well-known consistency property. Thus, if Θ is the set of possible states of nature, and we wish to test the composite hypotheses

$$H_1 : \theta \in \Theta_1 \qquad \text{vs} \qquad H_2 : \theta \in \Theta_2 \tag{1.3}$$

where Θ_1 and Θ_2 are two disjoint finite subsets of Θ, then Chernoff proposed the following rule for choosing experiments: if n observations (n = 0,1,...) have been taken, denote the maximum-likelihood estimate of

θ, based on the n observations, by $\hat{\theta}_n$, and assume without loss of generality that $\hat{\theta}_n \in \Theta_1$. Then if experimentation is to continue, select the experiment $e \in \mathcal{E}$, a finite set of experiments, that maximises

$$I(\hat{\theta}_n) = \inf_{\phi \in \Theta_2} I(\hat{\theta}_n, \phi, e) \tag{1.4}$$

where e may be chosen by a probability distribution over $\mathcal{E}$ and

$$I(\theta,\phi,e) = \int \log\left\{\frac{f(x|\theta,e)}{f(x|\phi,e)}\right\} f(x|\theta,e)\,dx\,. \tag{1.5}$$

This rule is a maximin strategy in a two-person zero-sum game in which Nature selects ϕ from Θ_2 to minimise I, while the experimenter selects e from $\mathcal{E}$ to maximise I, and $I(\hat{\theta}_n, \phi, e)$ is the payoff matrix. Chernoff showed that, under quite general conditions, this procedure is asymptotically optimal in the sense that, for each $\theta \in \Theta$, the risk $R(\theta)$ satisfies

$$R(\theta) \approx \frac{-c \log c}{I(\theta)} \qquad \text{as} \qquad c \to 0 \tag{1.6}$$

and that if any other procedure does better for some value of θ, it must do worse by an order of magnitude for some other value of θ.

Consequently, Bessler (1960) extended Chernoff's results to the case where $\mathcal{E}$ may be infinite and the two hypotheses could be replaced by some finite number of hypotheses, and considered the performance of this procedure in several relevant examples. (Some of Bessler's results will be considered in Chapter 4). Then Albert (1961) extended the results to the case where Θ_1 and Θ_2 may both be infinite sets. At this stage, however, it was found in many cases that Θ_1 and Θ_2 have a common boundary, and that $I(\theta)$ is zero on the boundary. Investigation of this problem (see Schwarz (1962)) indicated that the assumption of an indifference zone would solve this problem; that is, Θ_1 and Θ_2 may be "separated" by a third region Θ_3, the indifference region, such that if $\theta \in \Theta_3$, then the cost of accepting either hypothesis is zero. Thus Kiefer and Sacks (1963) were able to show that Chernoff's procedure was asymptotically optimal in the general case of k infinite hypotheses $\Theta_1, \Theta_2, \ldots, \Theta_k$ separated by indifference zones.

The assumption of an indifference zone does restrict the class of problems to which Chernoff's procedure can be applied, but the major shortcoming in all the situations described above is of course the performance of this procedure for small or moderate sample sizes. Whilst large sample considerations are unquestionably of value in gaining insight into the problem, and asymptotic optimality is certainly a desirable property, it still does not follow that Chernoff's rule is effective in problems where the number of observations is bounded or likely to be small. In fact, it has been found that this rule selects relatively uninformative experiments in some small sample size examples. Of course, this is hardly surprising, for as Chernoff himself points out, his procedure "..... treats the estimate of θ based on a few observations with as much respect as that based on many observations." (Chernoff (1975)). All the advantages which were available under large sample situations are no longer present: the maximum-likelihood estimate may be very different from the true value θ, the performance characteristics in the original problem depend strongly on the prior probabilities and the costs of incorrect terminal decisions, and hence the efficiency of any procedure based solely on information numbers is likely to be poor.

It is the purpose of this thesis to reconsider the problem of the sequential design of experiments from a Bayesian viewpoint. Although not universally acceptable on philosophical grounds, the Bayesian approach does have a number of distinct advantages in this problem: firstly, the optimal solution is available (at least in principle); secondly, posterior probability is an effective way of combining the experimenter's initial knowledge and his increasing units of information; and thirdly, a prior probability distribution over the possible states of nature provides a reasonable criterion for deciding which experiments are more informative than others before any data has been observed. Indeed, quoting Chernoff

again, "The use of the posterior probabilities in selecting (the next experiment) may add somewhat to the difficulty in implementing a procedure but seems quite sensible in attacking this problem." (Chernoff (1972)). Moreover, the only rule proposed so far as effective in the small sample case, that of Blot and Meeter (1973), is based on posterior probabilities. For a finite space $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$ they propose that, after n observations (n = 0,1,2,...), the next experiment chosen is that $e \in \mathcal{E}$ which maximises

$$\sum_{j=1}^{k} p_{nj} \, I(\hat{\theta}_n, \theta_j, e) \tag{1.7}$$

where p_{nj} denotes the posterior probability of θ_j, given n independent observations, and $\mathcal{E}$ is again finite. This is a more reasonable criterion than the pessimistic maximin criterion of Chernoff, as it modifies "Nature's strategy" or likely strategy according to the observations taken so far. However, it is still based on the maximum likelihood estimate, which as we have already mentioned is not necessarily a good estimate of θ when the sample size is small.

In Chapter 2 the results of Haggstrom (1966) are re-derived in terms of the Bayes statistical decision theory model. Defining a sequential design procedure to be a triple $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$, where $\underline{\phi}$ is a stopping rule, $\underline{\gamma}$ is an experiment rule, and $\underline{\delta}$ is a terminal decision rule, it is shown that an optimal sequential design procedure exists in each of the following situations: (a) the number of observations available is bounded by some integer N, and (b) the number of observations available is unbounded. Here optimality is defined in terms of minimising the Bayes risk for the problem. Further, it is shown that the Bayes risk for the problem in (a) converges to the Bayes risk for the problem in (b) as the upper bound N increases. Because the solution of the optimal procedure in (a) is a constructive one, it is thus possible to construct "near-optimal" solutions

to the general, i.e. unbounded, problem. Finally, it is shown that the experimenter can restrict himself to a certain class of experimental design procedures in his search for an optimal procedure; that is, there exists an essentially complete class of procedures for the sequential design problem.

In Chapter 3, two examples of the optimal sequential design procedure are considered. The first is an identification problem concerning the parameters of several binomial populations, and the second is a ranking problem concerning the means of two normal populations. The second example illustrates the difficulties involved in computing the optimal procedure for non-trivial problems.

Finally, Chapter 4 is devoted to the study of a myopic or One-Step-Ahead experiment rule, that is, a rule which selects the next experiment as if it were the last experiment to be performed. For a modified goal, it is shown that this myopic rule ensures convergence of the posterior probability of the true parameter value to one in examples considered by Bessler, and that it guarantees a certain optimality criterion. Although it is shown that a myopic rule will not be optimal in general, the performance of the myopic rule in simulations of small sample size problems is observed to be as good or better than that of Bessler's rule. The examples studied are identification problems concerning the parameters of several binomial populations, and two normal populations.

3. TWO RELATED PROBLEMS

It is perhaps worthwhile to mention two problems involving the sequential design of experiments. Although not considered in this thesis, some of the proofs and methods used in these areas of research are similar to those employed in Chapter 4. Also, effective solutions in each of these problems are of great practical value, thus resulting in a great deal of research in recent years.

(1) The Two-Armed Bandit Problem

First proposed in Robbins (1952), this problem relates to the choice of which slot machine to play in order to maximise winnings. In its simplest form, one is given the choice of two slot machines to play. For each pull on the arm of machine i, there is a constant probability of a payoff, p_i (i = 1,2). There are two possible simple hypotheses:

$$H_1 : (p_1,p_2) = (\nu_1,\nu_2) \qquad \text{vs} \qquad H_2 : (p_1,p_2) = (\nu_2,\nu_1)$$

where ν_1 and ν_2 are two known distinct probabilities. This corresponds to an identification problem mentioned in the preceding section; the probabilities are known, but the association with the machines is not. (See Chapter 3 for the definition of an identification problem). After each observation or pull, it is possible to stay on the current machine or switch to the other, until N observations have been made. Assuming the payoff on each machine is the same, what experimental design rule maximises the expected total payoff after N plays?

As a solution to this problem, Robbins proposed the Play-The-Winner Rule: select the first play at random, and thereafter continue on the current machine if a success is obtained, and switch to the other machine if a failure is obtained. Although this rule seems reasonable, it is not optimal. The optimal rule, discovered by Feldman (1962), is; play machine i if $P(p_i = \max\{\nu_1,\nu_2\})$ is at least $\frac{1}{2}$, where $P(\cdot)$ denotes the posterior probability of the hypotheses at the current stage of experimentation.

From a Bayesian point of view, the above example corresponds to a 2-point prior probability distribution on (p_1,p_2): p_1 and p_2 are random variables (r.v.s) with $P((p_1,p_2) = (\nu_1,\nu_2)) = g$ and $P((p_1,p_2) = (\nu_2,\nu_1)) = 1 - g$ for some g, $0 \leqslant g \leqslant 1$. Thus p_1 and p_2 are dependent r.v.s. For a general probability distribution of (p_1,p_2) over the range $[0,1]^2$, the

optimal procedure is not known. However, if p_1 and p_2 are independent, Berry (1972) has shown that the optimal procedure will continue on the same machine as long as a success occurs, but it may either stay or switch if a failure occurs. In general, the optimal procedure is found by backward induction and is extremely difficult to implement. Consequently, Berry (1978) proposed a myopic rule based on Feldman's rule as an approximate solution to the general problem and showed that its performance is quite good when compared to the optimal procedure. This thesis discusses a similar problem and proposes a similar solution.

(2) Sequential Medical Trials

A medical trial is defined to be an experiment to choose the better of two or more treatments for the same disease. Thus a sequential medical trial involves administering the competing treatments on successive patients until one of the treatments is chosen to be superior. Without an experimental design, however, a large number of patients will be administered inferior treatment. Consequently, the physician faces an ethical dilemma: he is bound to give each patient the best treatment (in his opinion), but he cannot decide which is the best treatment without observing responses from all of them. Although an experimental design procedure will not provide the whole answer, a search for a "best" procedure is certainly justified. In recent years many solutions have been proposed. Some assume the existence of the so-called finite patient horizon: there are N patients to be treated altogether. Consequently, two-stage procedures have been proposed, in which n of the N patients are treated with the different treatments, and on the basis of that sample, the remaining N-n patients are given the treatment found to be best. Other rules have been proposed which minimise the expected number of patients who are given inferior treatments. Note that large sample results are of little value here; the physician wishes to reach a decision as soon as possible.

A review of much of the recent work devoted to sequential medical trials may be found in Hoel, Sobel, and Weiss (1975).

CHAPTER II

CONSTRUCTION OF THE BAYES SEQUENTIAL DESIGN PROCEDURE

The results in this chapter were first derived by Haggstrom (1966) as the optimal solution of a sequential game in which the experimenter, stopping after a series of experiments denoted by a, receives a payoff Za. By considering the maximal submartingale relative to the stochastic process $\{Za, Fa, a \in A\}$, Haggstrom characterised the optimal solution of the game in both the truncated and nontruncated cases, and showed that the solution in the nontruncated case is the appropriate limit of solutions in the truncated case. These results are derived below as solutions to the obvious generalisation of a statistical sequential decision problem with a choice of experiments, and may be seen as corresponding to the results of Arrow, Blackwell and Girshick (1949) for the sequential decision problem. Furthermore, Haggstrom's results are extended to the case where the space of available experiments, $\mathcal{E}$, is infinite.

1. THE MODEL

(1) Description Of The Problem

The process considered is one in which experimentation is conducted at a potentially unbounded number of stages $j = 0,1,2,\ldots$. At each stage j, the statistician must decide whether to continue sampling or to stop. If he decides to continue sampling, he must then select the experiment E_{j+1} to be performed at the (j+1)-th stage. This selection is determined by a randomised experiment rule. At each stage j, the performance of an experiment E_j results in the observation of a random variable X_j, whose distribution depends on the past observations $X_1,\ldots, X_{j-1}$, past experiment choices $E_1,\ldots, E_j$, and an unknown state

of nature θ.

If the statistician decides to stop sampling, however, he must then take a terminal action A, chosen according to a randomised terminal decision rule. Upon doing so, and observing the value A = a, he incurs a real-valued loss $L(\theta, a)$. If stopping occurs after the j^{th} stage of experimentation, $j = 0,1,\ldots$, the statistician also incurs a real-valued sampling cost $c_j(\theta, \underline{x}^j, \underline{e}^j)$, whose value depends on the unknown state of nature θ, the past observation values $\underline{x}^j = (x_1,\ldots, x_j)$, and the past experiment choices $\underline{e}^j = (e_1,\ldots, e_j)$. Thus the total loss incurred is

$$L(\theta, a) + c_j(\theta, \underline{x}^j, \underline{e}^j). \tag{2.1}$$

The decision whether to stop or continue sampling is made on the basis of a randomised stopping rule, which determines the distribution of the random number of observations N.

It is henceforth assumed that $\theta \in \Theta$, X_j takes values $x_j \in \mathcal{X}$, E_j takes values $e_j \in \mathcal{E}$, and the terminal decision r.v. A takes values $a \in \mathcal{A}$, where the sets Θ, $\mathcal{X}$, $\mathcal{E}$ and $\mathcal{A}$ are all Borel sets in Euclidean spaces. Also, it is convenient to denote the absence of any observations by $(\underline{X}^0, \underline{E}^0)$. Note that the convention that lower case letters denote the values realised by the corresponding upper case r.v.s. is employed.

Informally, the probability structure of the problem can be defined as follows: the sequence of population distributions $\mathrm{Dist}_\theta(X_{j+1} | \underline{X}^j, \underline{E}^{j+1})$, $j = 0,1,\ldots$, is implied in the nature of the problem. Each such distribution describes the probability distribution of X_{j+1} given the past observations $\underline{x}^j$ and the past experiment selections $\underline{e}^{j+1}$. The stopping, selection and terminal decision rules are chosen by the statistician. The stopping rule $\underline{\phi} = (\phi_0, \phi_1,\ldots)$ is a sequence of functions $\phi_j(\underline{x}^j, \underline{e}^j)$ representing the probability that experimentation ceases after the j^{th} stage and observations $\underline{x}^j$ and experiments $\underline{e}^j$ have been observed, and is

assumed to be independent of θ and the terminal decision rule. The selection rule $\underline{\gamma} = (\gamma_0, \gamma_1, \ldots)$ corresponds to the sequence of conditional distributions $\text{Dist}(E_{j+1} | \underline{X}^j, \underline{E}^j)$, each of which is determined by the conditional density γ_j and is assumed to be independent of θ and the terminal decision rule. Finally, the terminal decision rule $\underline{\delta} = (\delta_0, \delta_1, \ldots)$ corresponds to the sequence of conditional distributions $\text{Dist}(A | \underline{X}^j, \underline{E}^j)$, each of which is determined by the conditional density δ_j and is independent of θ.

This probability structure is now formalised for the case where the random variables have Euclidean domains and ranges. Firstly, define the σ-finite measure spaces

$$(X, \mathcal{B}_X, \mu_X) \quad \text{and} \quad (\mathcal{E}, \mathcal{B}_E, \mu_E) \tag{2.2}$$

corresponding to observation and experiment r.v.s, respectively. There is also defined the σ-finite measure space

$$(\Theta, \mathcal{B}_\theta, \mu_G) \tag{2.3}$$

where μ_G is a measure on Θ induced by the prior distribution $G(\theta)$ of the state of nature θ. (Note that θ is used to denote both a random variable and an observed value). Finally, the σ-finite measure spaces

$$(M, \mathcal{B}(N), \mu_N) \quad \text{and} \quad (A, \mathcal{B}_A, \mu_A) \tag{2.4}$$

are defined, where μ_N is a counting measure on the non-negative integers $M = \{0,1, \ldots,\}$ corresponding to the random variable N, the stopping time, and A is the space of terminal decisions.

The basic sample space is thus

$$\Omega = \Theta \times X^\infty \times \mathcal{E}^\infty \times M \times A . \tag{2.5}$$

It is convenient to introduce the sub σ-fields (of sets in Ω)

$$F_n = \mathcal{B}(\underline{X}^n, \underline{E}^n), \qquad G_{n-1} = \mathcal{B}(\underline{X}^{n-1}, \underline{E}^n) \tag{2.6}$$

for n = 1,2, ... These σ-fields satisfy the nesting relationship

$$G_0 \subset F_1 \subset \ldots \subset G_{n-1} \subset F_n \subset \ldots \ . \tag{2.7}$$

Thus, for example, $F_n = \mathcal{B}(\underline{X}^n, \underline{E}^n)$ is the Borel field generated by sets of the form

$$\Theta \times B_1 \times C_1 \times \ldots \times B_n \times C_n \left(\prod_{j=n+1}^{\infty} X\right) \left(\prod_{j=n+1}^{\infty} \&\right) \times M \times A \tag{2.8}$$

where $B_j \in \mathcal{B}(X)$ and $C_j \in \mathcal{B}(E)$ for j = 1,2, ... n.

Define

$$F = \mathcal{B}(\underline{X}, \underline{E}) \tag{2.9}$$

to be the minimal σ-field containing the sequence of σ-fields (2.7), where $\underline{X} = (X_1, X_2, \ldots)$ and $\underline{E} = (E_1, E_2, \ldots)$ denote infinite sequences of observation and experiment r.v.s, respectively. Finally, the σ-fields

$$\begin{aligned} S &= \mathcal{B}(\underline{X}, \underline{E}, N), \quad T = \mathcal{B}(\underline{X}, \underline{E}, N, A) \\ U &= \mathcal{B}(\theta, \underline{X}, \underline{E}, N, A) \end{aligned} \tag{2.10}$$

are defined, which satisfy the nesting relationship

$$F \subset S \subset T \subset U \tag{2.11}$$

(2) Density Functions: The Components Of The Sequential Design Rule

(a) Population Densities. It is assumed that for each $\theta \in \Theta$ there is a sequence of population densities $\{f_\theta(x_{j+1}|\underline{x}^j, \underline{e}^{j+1})\}$. Each conditional density $f_\theta(x_{j+1}|\underline{x}^j, \underline{e}^{j+1})$ is non-negative, F_{j+1} - measurable, and satisfies

$$\int f_\theta(x_{j+1}|\underline{x}^j, \underline{e}^{j+1}) d\mu_X(x_{j+1}) = 1 \qquad \text{(a.s. } G_j\text{)}. \tag{2.12}$$

Frequently each density $f_\theta(x_{j+1}|\underline{x}^j, \underline{e}^{j+1})$ is of the form $f_\theta(x_{j+1}|e_{j+1})$,

i.e. the distribution of X_{j+1}, given θ, depends on the past observations $\underline{x}^j$ and the past experiments $\underline{e}^{j+1}$ only through the last experiment e_{j+1}, which generates the r.v. X_{j+1}. This condition is a generalisation of the "i.i.d." condition in decision theory.

The class of population densities is determined in advance. The statistician's task is to determine the stopping, experiment, and terminal decision rules according to some optimality criterion.

(b) The Stopping Rule. $\underline{\phi} = (\phi_0, \phi_1, \ldots)$ consists of F_j-measurable functions $\phi_j(\underline{x}^j, \underline{e}^j)$, each of which satisfies $0 \leqslant \phi_j \leqslant 1$.

(c) The Experiment Rule. $\underline{\gamma} = (\gamma_0, \gamma_1, \ldots)$ consists of G_j-measurable, non-negative, conditional densities $\gamma_j(e_{j+1}|\underline{x}^j, \underline{e}^j)$, each of which satisfies

$$\int \gamma_j(e_{j+1}|\underline{x}^j, \underline{e}^j)\, d\mu_E(e_{j+1}) = 1 \qquad (\text{a.s. } F_j) \tag{2.13}$$

(d) The Terminal Decision Rule. $\underline{\delta} = (\delta_0, \delta_1, \ldots)$ consists of non-negative, $\mathcal{T}$-measurable densities $\delta_j(a|\underline{x}^j, \underline{e}^j)$ satisfying

$$\int \delta_j(a|\underline{x}^j, \underline{e}^j)\, d\mu_A(a) = 1 \qquad (\text{a.s. } F_j) \tag{2.14}$$

for $j = 0, 1, \ldots$.

(e) A Sequential Design Procedure is a triple $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$, where $\underline{\phi}$ is a stopping rule, $\underline{\gamma}$ is an experiment rule, and $\underline{\delta}$ is a terminal decision rule.

(3) Definition Of Probability Measures

Appropriate probability measures on the σ-fields G_0, F_1, ..., G_{n-1}, F_n, ..., F, S, $\mathcal{T}$, and U are defined using a method of Doob's described in Gray (1968). Firstly, define

$$dP_{\gamma_0}(e_1) = \gamma_0(e_1)\, d\mu_E(e_1) \tag{2.15}$$

on G_0, and

$$dP_{f_\theta,\gamma_0}(x_1,e_1) = f_\theta(x_1|e_1)\, dP_{\gamma_0}(e_1)\, d\mu_X(x_1) \tag{2.16}$$

on F_1. Proceeding inductively, given the measure $P_{f_\theta,\underline{\gamma}^{n-1}}$ on F_n, define the measure

$$dP_{f_\theta,\underline{\gamma}^{n}}(\underline{x}^n,\underline{e}^{n+1}) = \gamma_n(e_{n+1}|\underline{x}^n,\underline{e}^n)\, dP_{f_\theta,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n)\, d\mu_E(e_{n+1}) \tag{2.17}$$

on G_n, and the measure

$$dP_{f_\theta,\underline{\gamma}^{n}}(\underline{x}^{n+1},\underline{e}^{n+1}) = f_\theta(x_{n+1}|\underline{x}^n,\underline{e}^{n+1})\, dP_{f_\theta,\underline{\gamma}^{n}}(\underline{x}^n,\underline{e}^{n+1})\, d\mu_X(x_{n+1}) \tag{2.18}$$

on F_{n+1}, where $\underline{\gamma}^{n-1} = (\gamma_0, \gamma_1, \ldots \gamma_{n-1})$ denotes the vector of the first n components of the experiment rule $\underline{\gamma}$. Thus, for n = 1,2,..., the probability spaces

$$(\Omega, G_{n-1}, P_{f_\theta,\underline{\gamma}^{n-1}}) \quad , \quad (\Omega, F_n, P_{f_\theta,\underline{\gamma}^{n-1}}) \tag{2.19}$$

are defined. By the Kolmogorov extension theorem, there is a unique probability space

$$(\Omega, F, P_{f_\theta,\underline{\gamma}}) \tag{2.20}$$

induced by the sequence of probability spaces (2.19). Further, defining

$$\psi_n(\underline{x}^n,\underline{e}^n) = \phi_n(\underline{x}^n,\underline{e}^n) \prod_{j=0}^{n-1} (1 - \phi_j(\underline{x}^j,\underline{e}^j)) \tag{2.21}$$

to denote the conditional probability $P\{N = n \mid \underline{x}, \underline{e}, N \geqslant n\}$, it follows that the probability space

$$(\Omega, S, P_{f_\theta,\underline{\phi},\underline{\gamma}}) \tag{2.22}$$

is defined by

$$dP_{f_\theta,\underline{\phi},\underline{\gamma}}(\underline{x},\underline{e},n) = \psi_n(\underline{x}^n,\underline{e}^n)\, dP_{f_\theta,\underline{\gamma}}(\underline{x},\underline{e})\, d\mu_N(n) \tag{2.23}$$

for $n = 0,1,\ldots$. By employing the terminal decision rule $\underline{\delta}$, the probability space

$$(\Omega, \mathcal{T}, P_{f_\theta,\underline{\phi},\underline{\gamma},\underline{\delta}}) \tag{2.24}$$

is defined by

$$dP_{f_\theta,\underline{\phi},\underline{\gamma},\underline{\delta}}(\underline{x},\underline{e},n,a) = \delta_n(a|\underline{x}^n,\underline{e}^n)\,dP_{f_\theta,\underline{\phi},\underline{\gamma}}(\underline{x},\underline{e},n)\,d\mu_A(a) \ . \tag{2.25}$$

Before defining the last measure, it is noted that the measure (2.25) enables computation of the risk, given θ, of using the sequential design procedure $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$. If it is assumed that the terminal loss $L(\theta,a)$ is $\mathcal{B}(A)$-measurable for all θ and that the sampling cost $C_n(\theta, \underline{x}^n, \underline{e}^n)$ is $\mathcal{F}_n$-measurable for each n and all θ, it follows that the total loss

$$L(\theta,A) + C_N(\theta, \underline{X}^N, \underline{E}^N) \tag{2.26}$$

is $\mathcal{T}$-measurable. If it is also assumed that both the terminal loss and the sampling cost are simultaneously bounded (usually bounded below), then the <u>risk</u>

$$R(\theta,(\underline{\phi},\underline{\gamma},\underline{\delta})) = E_{f_\theta,\underline{\phi},\underline{\gamma},\underline{\delta}}L(\theta,A) + E_{f_\theta,\underline{\phi},\underline{\gamma}}C_N(\theta,\underline{X}^N,\underline{E}^N) \tag{2.27}$$

is well-defined for all θ, and all sequential design procedures $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$. Finally, the probability space

$$(\Omega, \mathcal{U}, P_{G,f_\theta,\underline{\phi},\underline{\gamma},\underline{\delta}}) \tag{2.28}$$

is defined by the measure

$$dP_{G,f_\theta,\underline{\phi},\underline{\gamma},\underline{\delta}}(\theta,\underline{x},\underline{e},n,a) = dP_{f_\theta,\underline{\phi},\underline{\gamma},\underline{\delta}}(\underline{x},\underline{e},n,a)\,d\mu_G(\theta) \ . \tag{2.29}$$

If it is also assumed that $L(\theta,a)$ is $\mathcal{B}(\theta)$-measurable for all a, and $C_n(\theta, \underline{x}^n, \underline{e}^n)$ is $\mathcal{B}(\theta)$-measurable for each n and all $(\underline{x}^n, \underline{e}^n)$, then the total loss (2.26) is also $\mathcal{U}$-measurable, and the <u>Bayes risk</u>

$$r(G,(\underline{\phi},\underline{\gamma},\underline{\delta})) = E_{G,f_\theta,\underline{\phi},\underline{\gamma},\underline{\delta}} L(\theta,A) + E_{G,f_\theta,\underline{\phi},\underline{\gamma}} C_N(\theta,\underline{X}^N,\underline{E}^N) \tag{2.30}$$

is well-defined for all prior distributions G and all procedures $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$.

Writing the non-negative, $\mathcal{F}_n$-measurable density function associated with the measure $P_{f_\theta,\underline{\gamma}^{n-1}}$ as

$$f_{\theta,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n) = \prod_{j=1}^{n} f_\theta(x_j|\underline{x}^{j-1},\underline{e}^j)\gamma_{j-1}(e_j|\underline{x}^{j-1},\underline{e}^{j-1}) \tag{2.31}$$

$n = 1,2,\ldots$, then the Bayes risk (2.30) may be written as

$$\begin{aligned} r(G,(\underline{\phi},\underline{\gamma},\underline{\delta})) &= E_G\{R(\theta,(\underline{\phi},\underline{\gamma},\underline{\delta}))\} \\ &= \int R(\theta,(\underline{\phi},\underline{\gamma},\underline{\delta}))\,d\mu_G(\theta) \\ &= \sum_{n=1}^{\infty} \int_{\mathcal{X}^n} \int_{\mathcal{E}^n} \int_{\Theta} \psi_n(\underline{x}^n,\underline{e}^n)\,[L(\theta,\delta_n(\underline{x}^n,\underline{e}^n)) + C_n(\theta,\underline{x}^n,\underline{e}^n)]\,. \\ &\qquad f_{\theta,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n)\,d\mu_G(\theta)\,d\mu_E(e_1)\ldots d\mu_E(e_n)\,d\mu_X(x_1)\ldots d\mu_X(x_n) \end{aligned} \tag{2.32}$$

where $\delta_n(\underline{x}^n, \underline{e}^n)$ is defined implicitly by

$$L(\theta,\delta_n(\underline{x}^n,\underline{e}^n)) = \int_A L(\theta,a)\,\delta_n(a|\underline{x}^n,\underline{e}^n)\,d\mu_A(a) \tag{2.33}$$

and the order of integration has been rearranged using the Fubini-Tonnelli theorem (see e.g. Kingman and Taylor (1966), p.144), the appropriate measurability condition being assumed above. This may be compared with the expression for the Bayes risk in the usual sequential decision problem (see, e.g. Ferguson (1967) pp.313-314, equations (7.7) and (7.8), or deGroot (1970) p.275, equation (1)).

2. BAYES SEQUENTIAL DESIGN PROCEDURE FOR THE TRUNCATED PROBLEM

The choice of the sequential design procedure by the statistician is made on the basis of some optimality criterion; that is, the procedure chosen is better in some sense than any other procedure on average. The

optimality criterion used in this chapter is that of minimising the Bayes risk. Thus, the sequential design procedure $(\underline{\phi}^*, \underline{\gamma}^*, \underline{\delta}^*)$ is said to be Bayes for the sequential design problem with respect to the prior distribution G if

$$r(G,(\underline{\phi}^*,\underline{\gamma}^*,\underline{\delta}^*)) \leqslant r(G,(\underline{\phi},\underline{\gamma},\underline{\delta})) \quad \text{for all sequential design procedures } (\underline{\phi},\underline{\gamma},\underline{\delta}) \tag{2.34}$$

It is said to be ε-Bayes if, for some $\varepsilon > 0$,

$$r(G,(\underline{\phi}^*,\underline{\gamma}^*,\underline{\delta}^*)) \leqslant r(G,(\underline{\phi},\underline{\gamma},\underline{\delta})) + \varepsilon \quad \text{for all sequential design procedures } (\underline{\phi},\underline{\gamma},\underline{\delta}) \tag{2.35}$$

In general, if no Bayes procedure exists, an ε-Bayes procedure may be considered optimal. In this section, it is shown that a Bayes, or possibly ε-Bayes, procedure exists for the problem truncated after J observations, i.e. the class of stopping rules is restricted to those rules $\underline{\phi}$ that satisfy

$$\phi_J(\underline{x}^J, \underline{e}^J) \equiv 1 \quad \text{for all } (\underline{x}^J, \underline{e}^J) \in \mathcal{X}^J \times \mathcal{E}^J . \tag{2.36}$$

In this situation the statistician is permitted to stop after any stage of experimentation 0,1, ... up to the J^{th} stage, but he must do so no later than the J^{th} stage.

The existence of a Bayes sequential design procedure for the truncated problem with respect to some prior distribution G is shown in three stages: firstly, given any stopping rule $\underline{\phi}$ and any experiment rule $\underline{\gamma}$, there is a terminal decision rule $\underline{\delta}^*$ such that the risk of procedure $(\underline{\phi}, \underline{\gamma}, \underline{\delta}^*)$ is no greater than that of procedure $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$ for any terminal decision rule $\underline{\delta}$; secondly, given any experiment rule $\underline{\gamma}$, there is a stopping rule $\underline{\phi}^*$ (whose structure depends on $\underline{\gamma}$) such that the risk of $(\underline{\phi}^*, \underline{\gamma}, \underline{\delta}^*)$ is no greater than that of $(\underline{\phi}, \underline{\gamma}, \underline{\delta}^*)$ for any stopping rule $\underline{\phi}$,

where $\underline{\phi}$ and $\underline{\phi}^*$ both satisfy the truncation condition (2.36); and thirdly, there is an experiment rule $\underline{\gamma}^*$ (whose structure depends on $\underline{\phi}^*$) such that the risk of $(\underline{\phi}^*, \underline{\gamma}^*, \underline{\delta}^*)$ is no greater than that of $(\underline{\phi}^*, \underline{\gamma}, \underline{\delta}^*)$ for any experiment rule $\underline{\gamma}$. Moreover, the rules $\underline{\phi}^*$, $\underline{\gamma}^*$ and $\underline{\delta}^*$ are all defined constructively.

(1) The Optimal Terminal Decision Rule

It will be assumed henceforth that the prior distribution $G(\theta)$ possesses a density function $g(\theta)$: thus the basic probability space corresponding to (2.3) is defined by the measure $\nu(\theta)$, where

$$d\mu_G(\theta) = g(\theta)d\nu(\theta) . \tag{2.37}$$

Define the density function of $(\underline{X}^n, \underline{E}^n)$, unconditional on θ, by

$$f_{G,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n) = \int_\Theta f_{\theta,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n)g(\theta)d\nu(\theta) \tag{2.38}$$

and hence the posterior density of θ, given that $(\underline{X}^n, \underline{E}^n) = (\underline{x}^n, \underline{e}^n)$, by

$$g(\theta|\underline{x}^n,\underline{e}^n) = \frac{f_{\theta,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n)g(\theta)}{f_{G,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n)} \tag{2.39}$$

for $n = 1,2,\ldots$. Then the Bayes decision rule $\delta_n^*(\underline{x}^n, \underline{e}^n)$ for the fixed sample size problem based on $(\underline{x}^n, \underline{e}^n)$, which minimises

$$\begin{aligned}
&\int_{\mathcal{X}^n}\int_{\mathcal{E}^n}\int_\Theta [L(\theta,\delta_n(\underline{x}^n,\underline{e}^n)) + c_n(\theta,\underline{x}^n,\underline{e}^n)] \\
&\quad \times f_{\theta,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n)g(\theta)d\nu(\theta) \prod_{i=1}^n d\mu(e_i) \prod_{j=1}^n d\mu(x_j)^* \\
&= \int_{\mathcal{X}^n}\int_{\mathcal{E}^n}\left[\int_\Theta \{L(\theta,\delta_n(\underline{x}^n,\underline{e}^n)) + c_n(\theta,\underline{x}^n,\underline{e}^n)\}g(\theta|\underline{x}^n,\underline{e}^n)d\nu(\theta)\right] \\
&\quad \times f_{G,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n) \prod_{i=1}^n d\mu(e_i) \prod_{j=1}^n d\mu(x_j)
\end{aligned} \tag{2.40}$$

is equivalent to the decision rule which minimises the expected posterior

*see footnote, next page

loss, i.e.

$$\int_{\Theta} L(\theta,\delta_n(\underline{x}^n, \underline{e}^n))g(\theta|\underline{x}^n, \underline{e}^n)d\nu(\theta) \tag{2.41}$$

for each $(\underline{x}^n, \underline{e}^n) \in \mathcal{X}^n \times \mathcal{E}^n$. Thus it may be taken as a definition of the Bayes decision rule that it is the rule which minimises the expected posterior loss. Moreover, the Bayes decision rule is independent of the experiment rule $\underline{\gamma}^{n-1}$.

Theorem 2.1. Let $\delta_n^*(\underline{x}^n, \underline{e}^n)$ be a Bayes decision rule for the fixed sample size problem based on $(\underline{X}^n, \underline{E}^n) = (\underline{x}^n, \underline{e}^n)$, chosen according to the experiment rule $\underline{\gamma}^{n-1} = (\gamma_0,\gamma_1,\ldots \gamma_{n-1})$. Then for any stopping rule $\underline{\phi}$ and any such experiment rule $\underline{\gamma} = (\gamma_0,\gamma_1,\ldots \gamma_{n-1}, \ldots)$ $r(G,(\underline{\phi}, \underline{\gamma}, \underline{\delta}))$ is minimised by $r(G, (\underline{\phi}, \underline{\gamma}, \underline{\delta}^*))$, where $\underline{\delta}^* = (\delta_0^*, \delta_1^*, \delta_2^*, \ldots)$.

Proof: From the definition of the Bayes risk $r(G, (\underline{\phi}, \underline{\gamma}, \underline{\delta}))$ in (2.32), it follows that (2.32) is minimised over $\underline{\delta}$ by choosing each component δ_n to minimise

$$\int_{\mathcal{X}^n}\int_{\mathcal{E}^n}\int_{\Theta} \psi_n(\underline{x}^n,\underline{e}^n)\,[L(\theta,\delta_n(\underline{x}^n,\underline{e}^n)) + c_n(\theta,\underline{x}^n,\underline{e}^n)]$$
$$\times f_{\theta,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n)g(\theta)d\nu(\theta) \prod_{i=1}^{n} d\mu(e_i) \prod_{i=1}^{n} d\mu(x_i) \tag{2.42}$$

, $n = 1,2, \ldots$. Since $\psi_n(\underline{x}^n, \underline{e}^n)$ is not a function of θ, however, this expression is also minimised by choosing each δ_n to minimise (2.41) for each $(\underline{x}^n, \underline{e}^n)$, $n = 1,2, \ldots$. □

Footnote: *Henceforth the suffices E and X are dropped from the appropriate measures.

Note: This result holds for all stopping rules $\underline{\phi}$, i.e. it holds for both the truncated and the nontruncated problem.

Remark: Although the result of Theorem 2.1 is hardly surprising, it is worth noting that in the experimental design situation, the optimal choice

of $\underline{\delta}$ is not only independent of $\underline{\phi}$ but also independent of $\underline{\gamma}$, the experiment rule. Roughly speaking, this is because $\underline{\delta}$ is independent of $\underline{\phi}$ and $\underline{\gamma}$ functionally and the choice of $\underline{\delta}^*$ depends on $(\underline{x}^n, \underline{e}^n)$ through the posterior density $g(\theta|\underline{x}^n, \underline{e}^n)$. From the defining equations (2.31), (2.38) and (2.39), however, it may be seen that $g(\theta|\underline{x}^n, \underline{e}^n)$ is independent of $\underline{\gamma}$, for any n.

(2) The Optimal Stopping Rule

In this section the notation of Ferguson (1967) Section 7.2, appropriately modified, is used to construct the optimal stopping rule $\underline{\phi}^{J*}$ for the problem truncated after J observations. Denote the minimum conditional Bayes expected loss, given that $(\underline{X}^j, \underline{E}^j) = (\underline{x}^j, \underline{e}^j)$, by

$$\rho_j(g_j) = \inf_{\delta_j} \int_\Theta L(\theta, \delta_j(\underline{x}^j, \underline{e}^j)) g(\theta|\underline{x}^j, \underline{e}^j) d\nu(\theta)$$

$$= \inf_{a \in A} \int_\Theta L(\theta, a) g(\theta|\underline{x}^j, \underline{e}^j) d\nu(\theta) \qquad (2.43)$$

$j = 0,1,\ldots$. It is assumed that each ρ_j is integrable. The minimum conditional expected loss plus cost of stopping after j observations $(\underline{x}^j, \underline{e}^j)$ is denoted by $U_j(g_j)$:

$$U_j(g_j) = \inf_{a \in A} \int_\Theta [L(\theta, a) + c_j(\theta, \underline{x}^j, \underline{e}^j)] g(\theta|\underline{x}^j, \underline{e}^j) d\nu(\theta)$$

$$= \rho_j(g_j) + \int_\Theta c_j(\theta, \underline{x}^j, \underline{e}^j) g(\theta|\underline{x}^j, \underline{e}^j) d\nu(\theta) \qquad (2.44)$$

Both U_j and ρ_j are functions of $(\underline{x}^j, \underline{e}^j)$, but it is notationally convenient not to show this. Similarly, the notation

$$g_{j-1}(X_j, E_j) \equiv g(\theta|\underline{x}^{j-1}, X_j, \underline{e}^{j-1}, E_j)$$

$$f_{\underline{\gamma}^{j-1}}(\underline{x}^j, \underline{e}^j|g) \equiv f_{G,\underline{\gamma}^{j-1}}(\underline{x}^j, \underline{e}^j)$$

$$f_{\gamma_{j-1}}(x_j, e_j|g_{j-1}) = \int_\Theta f_\theta(x_j|\underline{x}^{j-1}, \underline{e}^j) \gamma_{j-1}(e_j|\underline{x}^{j-1}, \underline{e}^{j-1})$$

$$\times g(\theta|\underline{x}^{j-1}, \underline{e}^{j-1}) d\nu(\theta)$$

$$= f_G(x_j|\underline{x}^{j-1},\underline{e}^j) \cdot \gamma_{j-1}(e_j|\underline{x}^{j-1},\underline{e}^{j-1}) \qquad (2.45)$$

is used, where $g_{j-1}(X_j,E_j)$ denotes the posterior probability density function conditional on $(\underline{X}^{j-1},\underline{E}^{j-1}) = (\underline{x}^{j-1},\underline{e}^{j-1})$ and the r.v.s. (X_j,E_j), $f_{\underline{\gamma}^{j-1}}(\underline{x}^j,\underline{e}^j|g)$ denotes the joint density function of $(\underline{x}^j,\underline{e}^j)$ unconditional on θ, defined by (2.38), and (similarly) $f_{\gamma_{j-1}}(x_j,e_j|g_{j-1})$ denotes the conditional density of (x_j,e_j) given $(\underline{X}^{j-1},\underline{E}^{j-1}) = (\underline{x}^{j-1},\underline{e}^{j-1})$, unconditional on θ. The second equality shows that this density is the product of two densities, the j^{th} component of the experiment rule $\underline{\gamma}$, and the conditional density of x_j given $(\underline{X}^{j-1},\underline{E}^j) = (\underline{x}^{j-1},\underline{e}^j)$, unconditional on θ. Viewed in this way, $f_{\gamma_{j-1}}(x_j,e_j|g_{j-1})$ is seen to correspond to the density $f_G(x_j|\underline{x}^{j-1})$ in the sequential decision problem (see e.g. equation (7.12), Ferguson (1967)), being the product of this density with the experiment density γ_{j-1}.

The optimal stopping rule $\underline{\phi}^{J*}$ for the sequential design problem truncated after J observations may now be described. Suppose the statistician has taken J - 1 observations $(\underline{X}^{J-1},\underline{E}^{J-1}) = (\underline{x}^{J-1},\underline{e}^{J-1})$. If he stops sampling without observing (X_J,E_J), his loss plus cost is $U_{J-1}(g_{J-1})$. If he observes (X_J,E_J), where E_J is chosen by the J-th component of his experiment rule γ_{J-1}, then his expected loss plus cost is

$$E_{\gamma_{J-1}}\{U_J[g_{J-1}(X_J,E_J)]\,|\,\mathcal{F}_{J-1}\} = \int_{\mathcal{X}}\int_{\mathcal{E}} U_J[g_{J-1}(x_J,e_J)] \times f_{\gamma_{J-1}}(x_J,e_J|g_{J-1})\,d\mu(e_J)\,d\mu(x_J) \qquad (2.46)$$

Thus the component ϕ_{J-1}^{J*} of the optimal Bayes stopping rule is

$$\phi_{J-1}^{J*}(\underline{x}^{J-1},\underline{e}^{J-1}) = \begin{cases} 1 & \text{if } U_{J-1}(g_{J-1}) < E_{\gamma_{J-1}}\{U_J[g_{J-1}(X_J,E_J)]\,|\,\mathcal{F}_{J-1}\} \\ \text{any} & \text{if } \quad = \\ 0 & \text{if } \quad > \end{cases} \qquad (2.47)$$

Therefore, the minimum conditional Bayes risk, given $(X^{J-1},E^{J-1}) = (\underline{x}^{J-1},\underline{e}^{J-1})$ and based on the last component γ_{J-1} of an experiment rule $\underline{\gamma}^{J-1} = (\gamma_0, \gamma_1, \ldots \gamma_{J-1})$ for the truncated problem, $V_{J-1}^{(\underline{\gamma}^{J-1})}$ is defined by

$$V_{J-1}^{(\underline{\gamma}^{J-1})}(g_{J-1}) = \min\{U_{J-1}(g_{J-1}), E_{\gamma_{J-1}}\{U_J[g_{J-1}(X_J,E_J)]|F_{J-1}\}\} \tag{2.48}$$

$V_{J-1}^{(\underline{\gamma}^{J-1})}(g_{J-1})$ may be thought of as the risk of the optimal continuation at the $(J-1)$-th stage, based on $\underline{\gamma}^{J-1}$, and is clearly F_{J-1}-measurable. Note that $V_{J-1}^{(\underline{\gamma}^{J-1})}$ is a function of $\underline{\gamma}^{J-1}$ only through its last component γ_{J-1}.

Moving back now to the $(J-2)$-th stage, the risk incurred from stopping and choosing a terminal decision based on $(\underline{X}^{J-2},\underline{E}^{J-2}) = (\underline{x}^{J-2},\underline{e}^{J-2})$, is $U_{J-2}(g_{J-2})$. If the statistician observes the r.v. $X_{J-1} = x_{J-1}$ generated by an experiment $E_{J-1} = e_{J-1}$, according to the second-to-last component γ_{J-2} of his experiment rule $\underline{\gamma}^{J-1}$, however, his expected risk is

$$E_{\gamma_{J-2}}\left\{V_{J-1}^{(\underline{\gamma}^{J-1})}[g_{J-2}(X_{J-1},E_{J-1})]|F_{J-2}\right\} \tag{2.49}$$

defined analagously to (2.46). The minimum conditional Bayes risk, given $(\underline{X}^{J-2},\underline{E}^{J-2}) = (\underline{x}^{J-2},\underline{e}^{J-2})$, and based on (the last two components of) the experiment rule $\underline{\gamma}^{J-1}$, is

$$V_{J-2}^{(\underline{\gamma}^{J-1})}(g_{J-2}) = \min\left\{U_{J-2}(g_{J-2}), E_{\gamma_{J-2}}\left\{V_{J-1}^{(\underline{\gamma}^{J-1})}[g_{J-2}(X_{J-1},E_{J-1})]|F_{J-2}\right\}\right\} \tag{2.50}$$

Proceeding inductively, the $V_j^{(\underline{\gamma}^{J-1})}$'s may be defined as $V_J^{(\underline{\gamma}^{J-1})} = U_J$ and

$$V_{j-1}^{(\underline{\gamma}^{J-1})}(g_{j-1}) = \min\left\{U_{j-1}(g_{j-1}), E_{\gamma_{j-1}}\left\{V_j^{(\underline{\gamma}^{J-1})}[g_{j-1}(X_j,E_j)]|F_{j-1}\right\}\right\} \tag{2.51}$$

for $j = J, J-1, \ldots, 1$. The components of the optimal stopping rule $\underline{\phi}^{J*} = (\phi_0^{J*}, \phi_1^{J*}, \ldots, \phi_J^{J*})$ are defined by

$$\phi_{j-1}^{J*}(\underline{x}^{j-1},\underline{e}^{j-1}) = \begin{cases} 1 & \text{if } U_{j-1}(g_{j-1}) \\ & \quad < E_{Y_{j-1}}\left\{V_j^{(\underline{\gamma}^{J-1})}[g_{j-1}(X_j,E_j)]\,|\,F_{j-1}\right\} \\ \text{any} & \text{if } = \\ 0 & \text{if } > \end{cases} \tag{2.52}$$

for $j = J, J-1, \ldots, 1$, and $\phi_J^{J*}(\underline{x}^J,\underline{e}^J) \equiv 1$. The following theorem is thus proved:

Theorem 2.2. Based on the prior density $g(\theta)$, $V_0^{(\underline{\gamma}^{J-1})}(g)$ is the total risk of the stopping rule $\underline{\phi}^{J*} = (\phi_0^{J*}, \phi_1^{J*}, \ldots, \phi_J^{J*})$ based on the experiment rule $\underline{\gamma}^{J-1}$, for the sequential design problem truncated after J observations. Furthermore, for $j = J-1, J-2, \ldots, 1$, having observed $(\underline{X}^j,\underline{E}^j) = (\underline{x}^j,\underline{e}^j)$, the risk from the optimal continuation is $V_j^{(\underline{\gamma}^{J-1})}(g_j)$, where g_j is the posterior density based on $(\underline{x}^j,\underline{e}^j)$.

Theorem 2.3. For any given sampling rule $\underline{\gamma}^{J-1}$ for the problem truncated after J observations, where J is some non-negative integer, the stopping rule $\underline{\phi}^{J*} = (\phi_0^{J*}, \phi_1^{J*}, \ldots, \phi_J^{J*})$ is optimal for the sequential design problem truncated after J observations with respect to the prior distribution G, in the sense that

$$r(G, (\underline{\phi}^{J*}, \underline{\gamma}^{J-1}, \underline{\delta}^{J*})) \leqslant r(G, (\underline{\phi}^{J}, \underline{\gamma}^{J-1}, \underline{\delta}^{J*})) \tag{2.53}$$

for all stopping rules $\underline{\phi}^J$ satisfying (2.36)

where $\underline{\delta}^{J*} = (\delta_0^*, \delta_1^*, \ldots, \delta_J^*)$ is a vector comprising the first $J+1$ components of $\underline{\delta}^*$.

Proof: This follows from Theorem 3.2 of Chow, Robbins and Siegmund (1971), p.50. □

Remark: Although the construction of the optimal stopping rule $\underline{\phi}^{J*}$ depends on the particular experiment rule $\underline{\gamma}^{J-1}$ employed by the statistician, it is nevertheless of the same form for any experiment rule. Therefore, if the choice of experimentation (possibly randomised) is determined in advance, the sequential rules $\underline{\delta}^{J*}$ and $\underline{\phi}^{J*}$ are optimal for the truncated problem. In particular, if $\underline{\gamma}^{J-1}$ specifies that the same experiment is to be performed at each stage, then $(\underline{\phi}^{J*}, \underline{\delta}^{J*})$ is the well-known Bayes sequential decision rule for the sequential decision problem truncated after J observations (see e.g. Ferguson (1967) Theorem 2, p.317).

(3) The Optimal Experiment Rule

In order to construct the optimal experiment rule $\underline{\gamma}^{(J-1)*}$ for the problem truncated after J observations, the following lemma is needed.

Lemma 2.1. Let $X(\in \mathcal{X})$ and $Y(\in \mathcal{Y})$ be, respectively, a random variable and a random vector, and denote $\mathcal{F}_X = \mathcal{B}(X)$ and $\mathcal{F} = \mathcal{B}(X, Y)$. Let $F(x)$ be any distribution function of X, and let $g(x,y)$ be any $\mathcal{F}$-measurable integrable function such that $g(x,y)$ is $\mathcal{F}_X$-measurable for any $y \in \mathcal{Y}$. Then for any $y \in \mathcal{Y}$, if

$$M_g(y) = \operatorname*{ess\,inf}_{x \in \mathcal{X}} \{g(x,y)\} \tag{2.54}$$

then

$$M_g(y) \leqslant \int_{\mathcal{X}} g(x,y)\, dF(x) \tag{2.55}$$

for all distribution functions F.

Proof: For any $y \in \mathcal{Y}$, the definition of $M_g(y)$ implies that

$$M_g(y) \leqslant g(x,y) \qquad (\text{a.s. } \mathcal{F}_X)$$

which implies

$$M_g(y) - g(x,y) \leqslant 0 \quad (\text{a.s. } \mathcal{F}_X)$$

and hence

$$\int_X (M_g(y) - g(x,y))\,dF(x) \leqslant 0$$

i.e.

$$M_g(y) \leqslant \int_X g(x,y)\,dF(x)$$

for any distribution function $F(x)$ of X. □

Let $\mathcal{E}^*$ be the set of all probability density functions of the r.v. E which are positive-valued on at most a countable subset of $\mathcal{E}$: if $\mathcal{E}$ is countable it may be chosen as the subset itself. Then it is shown below that an optimal Bayes (or ε-Bayes) experiment rule $\underline{\gamma}^{(J-1)*}$ exists for the problem truncated after J observations in the sense that

$$r(G, (\underline{\phi}^{J*}, \underline{\gamma}^{(J-1)*}, \underline{\delta}^{J*})) \leqslant r(G, (\underline{\phi}^{J*}, \underline{\gamma}^{J-1}, \underline{\delta}^{J*})) \tag{2.56}$$

for all experiment rules $\underline{\gamma}^{J-1} = (\gamma_0, \gamma_1, \ldots, \gamma_{J-1})$, each of whose components γ_j satisfies (2.13) and belongs to $\mathcal{E}^*$.

Let

$$\begin{aligned} M_{V_J^{(\underline{\gamma}^{J-1})}}(\underline{x}^{J-1}, \underline{e}^{J-1}) &= \operatorname*{ess\,inf}_{e_J \in \mathcal{E}} E\left\{V_J^{(\underline{\gamma}^{J-1})}[g_{J-1}(X_J, e_J)] \,|\, G_{J-1}\right\} \\ &= \operatorname*{ess\,inf}_{e_J \in \mathcal{E}} \int_X V_J^{(\underline{\gamma}^{J-1})}[g_{J-1}(x_J, e_J)] \\ &\qquad \times f_G(x_J | \underline{x}^{J-1}, \underline{e}^J)\,d\mu(x_J) \end{aligned} \tag{2.57}$$

where G_{J-1} was defined in (2.6), $f_G(x_J|\underline{x}^{J-1}, \underline{e}^J)$ was defined in (2.45), and $V_J^{(\underline{\gamma}^{J-1})} = U_J$ for all experiment rules $\underline{\gamma}^{J-1}$. Then by Lemma 2.1

$$M_{V_J^{(\underline{\gamma}^{J-1})}}(\underline{x}^{J-1}, \underline{e}^{J-1}) \leqslant E_{\gamma_{J-1}}\left\{V_J^{(\underline{\gamma}^{J-1})}[g_{J-1}(X_J, E_j)] \,|\, F_{J-1}\right\} \tag{2.58}$$

(where the RHS is the expected loss plus cost of observing (X_J, E_J) with

E_J chosen according to γ_{J-1} defined in (2.46)), for all experiment densities $\gamma_{J-1} \in \mathcal{E}^*$ satisfying (2.13). Since $L(\theta,a)$ and $C_J(\theta, \underline{x}^J, \underline{e}^J)$ are bounded below for all values of their respective arguments, it follows from Lemma 2.1 that $M_{V_J^{(\underline{\gamma}^{J-1})}}(\underline{x}^{J-1}, \underline{e}^{J-1}) \neq -\infty$ for all $(\underline{x}^{J-1}, \underline{e}^{J-1})$ and that either

$$E\left\{V_J^{(\underline{\gamma}^{J-1})}[g_{J-1}(X_J, e_J)] \,|\, G_{J-1}\right\} = M_{V_J^{(\underline{\gamma}^{J-1})}}(\underline{x}^{J-1}, \underline{e}^{J-1}) \quad \text{for at least one value of } e_J \in \mathcal{E} \tag{2.59}$$

or, for some $\varepsilon > 0$,

$$E\left\{V_J^{(\underline{\gamma}^{J-1})}[g_{J-1}(X_J, e_J)] \,|\, G_{J-1}\right\} \leq M_{V_J^{(\underline{\gamma}^{J-1})}}(\underline{x}^{J-1}, \underline{e}^{J-1}) + \varepsilon \quad \text{for at least one value of } e_J \in \mathcal{E}. \tag{2.60}$$

Thus, the last component of the optimal experiment rule $\underline{\gamma}^{(J-1)*}$ is the conditional density function (in fact a probability mass function) $\gamma_{J-1}^{(J-1)*}(e_J | \underline{x}^{J-1}, \underline{e}^{J-1})$ which assigns equal probability to each $e_J \in \mathcal{E}$ satisfying (2.59), or failing that, to each $e_J \in \mathcal{E}$ satisfying (2.60) for some $\varepsilon > 0$.

It follows that each component of $\underline{\gamma}^{(J-1)*}$ may be defined inductively in a fashion similar to the definition of $\underline{\phi}^{J*}$: given $\gamma_{J-1}^{(J-1)*}$, $\gamma_{J-2}^{(J-1)*}$, ..., $\gamma_j^{(J-1)*}$, define

$$M_{V_j^{(\underline{\gamma}^{J-1})}}(\underline{x}^{j-1}, \underline{e}^{j-1}) = \operatorname*{ess\,inf}_{e_j \in \mathcal{E}} E\left\{V_j^{(\underline{\gamma}^{J-1})}[g_{j-1}(X_j, e_j)] \,|\, G_{j-1}\right\} \tag{2.61}$$

for any experiment rule $\underline{\gamma}^{J-1}$ whose last $J-j$ components are $\gamma_j^{(J-1)*}, \ldots, \gamma_{J-1}^{(J-1)*}$, as $V_j^{(\underline{\gamma}^{J-1})}$ depends on $\underline{\gamma}^{J-1}$ only through its last $J-j$ components. Then by Lemma 2.1

$$M_{V_j^{(\underline{\gamma}^{J-1})}}(\underline{x}^{j-1},\underline{e}^{j-1}) \leq E_{\gamma_{j-1}}\left\{V_j^{(\underline{\gamma}^{J-1})}[g_{j-1}(X_j,E_j)]\,|\,F_{j-1}\right\} \tag{2.62}$$

(where the RHS appears in (2.51)) for all experiment densities $\gamma_{j-1} \in \mathcal{E}^*$ satisfying (2.13), and $M_{V_j^{(\underline{\gamma}^{J-1})}}(\underline{x}^{j-1},\underline{e}^{j-1}) \neq -\infty$ for all $(\underline{x}^{j-1},\underline{e}^{j-1})$ by virtue of Lemma 2.1 and the definition of $V_j^{(\underline{\gamma}^{J-1})}$. Then the j^{th} component of $\underline{\gamma}^{(J-1)*}$ is the conditional probability function $\gamma_{j-1}^{(J-1)*}(e_j|\underline{x}^{j-1},\underline{e}^{j-1})$ which assigns equal probability to each $e_j \in \mathcal{E}$ attaining the (essential) infimum in (2.61) (or "ε-attaining" it).

Theorem 2.4. If the (essential) infimum in (2.61) is attained for each $j = J, J-1,\ldots, 1$, then $\underline{\gamma}^{(J-1)*} = (\gamma_0^{(J-1)*}, \gamma_1^{(J-1)*},\ldots, \gamma_{J-1}^{(J-1)*})$ is the Bayes experiment rule for the problem truncated after J observations with respect to the prior distribution G in the sense that

$$r(G, (\underline{\phi}^{J*}, \underline{\gamma}^{(J-1)*}, \underline{\delta}^{J*})) \leq r(G, (\underline{\phi}^{J*}, \underline{\gamma}^{J-1}, \underline{\delta}^{J*})) \tag{2.63}$$

for all experiment rules $\underline{\gamma}^{J-1}$ for the truncated problem each of whose components γ_j satisfies (2.13) and belongs to $\mathcal{E}^*$. If at least one of the (essential) infimums is not attained, then $\underline{\gamma}^{(J-1)*}$ is ε-Bayes in the sense that

$$r(G, (\underline{\phi}^{J*}, \underline{\gamma}^{(J-1)*}, \underline{\delta}^{J*})) \leq r(G, (\underline{\phi}^{J*}, \underline{\gamma}^{J-1}, \underline{\delta}^{J*})) + \varepsilon\,. \tag{2.64}$$

In particular, (2.59) is true whenever $\mathcal{E}$ is finite.

Proof: This follows from Theorem 4.1 of Haggstrom (1966): in Appendix A1, his results are extended to the case when $\mathcal{E}$ is countable. □

Remarks:

1. Given a class of r.v.s. $\{y_t,\ t \in T\}$, it is well-known that $\operatorname{ess\,inf}_{t\in T} y_t$ may be taken as $\inf_{t\in C} y_t$ where C is some countable subset of T. Therefore, the countable subset of $\mathcal{E}$ on which $\mathcal{E}^*$ is defined may be taken

as $\bigcup_{j=1}^{J} \mathcal{E}_j$, where $\mathcal{E}_j$ is the countable subset of $\mathcal{E}$ along which the ess. inf. in (2.61) is attained; and hence $\gamma_j^{(J-1)*} \in \mathcal{E}_j^*$, $j = 0,1,\ldots, J-1$, is assured.

2. The definition of the components $\gamma_j^{(J-1)*}$ is not clear if the set of e_j's (ε-) attaining the ess. inf. in (2.61) is infinite. However, $\gamma_j^{(J-1)*}$ is defined to give each such e_j an equal probability of being chosen, i.e. it is unbiased in the sense that a fair die is unbiased. From the Bayesian point of view, however, any of the competing e_j's will do, and so a $\gamma_j^{(J-1)*}$ with positive mass on at most a finite set is acceptable.*

3. The interpretation of the optimal experiment rule $\underline{\gamma}^{(J-1)*}$ is the same as the well-known result concerning the Bayes terminal decision rule $\underline{\delta}^{J*}$; that the statistician may restrict himself to rules γ_j that are degenerate at a single point, provided that each such γ_j is G_j-measurable.

Theorems 2.1, 2.3 and 2.4 thus give rise to the principal result in this section:

Theorem 2.5. The sequential design procedure $(\underline{\phi}^{J*}, \underline{\gamma}^{(J-1)*}, \underline{\delta}^{J*})$ is (ε-) Bayes for the sequential design problem truncated after J observations with respect to the prior distribution G, in the sense that

$$r(G, (\underline{\phi}^{J*}, \underline{\gamma}^{(J-1)*}, \underline{\delta}^{J*}) \leqslant r(G, (\underline{\phi}^{J}, \underline{\gamma}^{J-1}, \underline{\delta}^{J})) \ (+\varepsilon) \tag{2.65}$$

for all sequential design procedures $(\underline{\phi}^{J}, \underline{\gamma}^{J-1}, \underline{\delta}^{J})$ based on the same sequential design problem (and each component γ_j of $\underline{\gamma}^{J-1}$ belongs to $\mathcal{E}_j^*$, $j = 0,1,\ldots, J-1$).

*The comments of deGroot (1970), pp.128-130, concerning decision rules are relevant here.

3. BAYES SEQUENTIAL DESIGN PROCEDURE FOR THE NON-TRUNCATED PROBLEM

If the restriction that at most J observations may be taken is removed from the problem, then it is no longer possible to generate the optimal stopping and experiment rules by the method of backward induction used in Section 2. (Recall that the terminal decision component $\underline{\delta}^*$ was shown to be Bayes for the non-truncated problem in Theorem 2.1). It must be shown that a) an optimal stopping rule and an optimal experiment rule exist in the non-truncated case, and, b) that the solution to the non-truncated problem may be approximated by the solution to the problem truncated after J observations, for successively larger values of J.

(1) Existence Of An Optimal Stopping Rule And An Optimal Experiment Rule For The Non-truncated Problem

The existence of an optimal stopping rule for the non-truncated problem for any fixed experiment rule $\underline{\gamma}$ may be shown in an analogous fashion to the proof for the non-truncated sequential decision problem; see Chow, Robbins and Siegmund (1971), Theorem 4.4, p.69. However, Haggstrom (1966), Theorem 4.3, shows the existence of both an optimal stopping rule and an optimal experiment rule for the non-truncated problem provided the sampling cost $C_n(\theta, \underline{x}^n, \underline{e}^n)$ satisfies

$$C_0 \leqslant C_1 \leqslant C_2 \leqslant \ldots, \quad \lim_{n\to\infty} C_n = +\infty \ , \quad \text{for all values of the arguments} \tag{2.66}$$

Again, Haggstrom's result is extended to the case where $\mathcal{E}$ is countable in Appendix A1.

Roughly speaking, an optimal stopping rule may fail to exist because, along the path of experimentation dictated by the experiment rule, the risk $U_j(g_j)$ decreases with j, and so it is optimal not to stop at any stage. As an infinite number of observations is undesirable (and

impossible!) in practice, it is usual to restrict the class of stopping rules $\underline{\phi}$ to those rules whose associated stopping times N (see (2.4) and (2.21)) are finite (a.s. S); and hence no rule can be optimal. If, however, the sampling cost increases monotonically to $+\infty$, it is possible to ensure that an optimal $\underline{\phi}$ exists. Moreover, since $C_n \to +\infty$ for all $(\theta, \underline{x}^n, \underline{e}^n)$, it follows that no experiment rule can give rise to an "infinite" stopping time, and so the existence of an optimal experiment rule is also guaranteed.

Consequently, for any experiment rule, the optimal stopping rule in the non-truncated problem will have a finite risk and give rise to a stopping time with finite expectation. These properties will be used in the next section.

(2) Approximating The Solution To The Non-truncated Problem By Solutions To The Truncated Problem

In this section it is shown that the solution to the non-truncated sequential design problem is the limit, as $J \to \infty$, of solutions to the sequential design problem truncated after J observations. Consequently, it is possible to construct a solution to the non-truncated problem which is arbitrarily close to the optimal solution.

Consider the stopping time N associated with a stopping rule $\underline{\phi}$ for the non-truncated problem. Then it follows from the definition of a stopping rule that, for any experiment rule $\underline{\gamma}$,

$$P_{f_\theta, \underline{\phi}, \underline{\gamma}}[N \in M,\ N < \infty] = 1 \tag{2.67}$$

and

$$\{N = n\} \in \mathcal{F}_n \quad \text{for } n = 0,1,\ldots \tag{2.68}$$

where $\{N = n\}$ is an abbreviation for $\{(\underline{x}^n, \underline{e}^n) : N = n\}$. Furthermore, the events $\{N \leqslant n\} = \bigcup_{i=0}^{n} \{N = i\}$ and $\{N > n\} = \{N \leqslant n\}^c$ are also $\mathcal{F}_n$-measurable

events, for each n. Note that these events depend on the sequence of population densities and experiment densities. In what follows, the experiment rule is taken to be the optimal rule $\underline{\gamma}^*$.

Given the optimal procedure $(\underline{\phi}^*, \underline{\gamma}^*, \underline{\delta}^*)$ for the general (non-truncated) problem, with associated stopping time N^*, and some positive integer n, let N_n^* be the truncated stopping time defined by

$$\{N_n^* = j\} = \{N^* = j\},\ j = 0,1,\ldots,\ n-1$$

$$\{N_n^* = n\} = \{N^* \geqslant n\} = \{N^* = n\} \cup \{N^* > n\} \tag{2.69}$$

i.e. N_n^* agrees with N^* through the first n - 1 observations, but it ensures that experimentation ceases after the n^{th} observation, if it has not ceased earlier, regardless of whether N^* specifies that further observations should be taken. The stopping rule $\underline{\phi}_n^*$ defined by N_n^* is clearly a stopping rule for the problem truncated after n observations, i.e. it satisfies (2.36) for J = n. Similarly, $\underline{\gamma}^*_{n-1} = (\gamma_0^*, \gamma_1^*,\ldots, \gamma^*_{n-1})$ denotes the vector of the first n components of $\underline{\gamma}^*$, and $\underline{\delta}_n^*$ (also written $\underline{\delta}^{n*}$ previously) denotes the vector of the first n + 1 components of $\underline{\delta}^*$.

Theorem 2.6. For n = 1,2,..., define

$$Q_n = \int_{\{N^* > n\}} \rho_n[g_n(\underline{x}^n, \underline{e}^n)]\, f_{G,\underline{\gamma}^*_{n-1}}(\underline{x}^n, \underline{e}^n) \prod_{i=1}^{n} d\mu(e_i) \prod_{j=1}^{n} d\mu(x_j) \tag{2.70}$$

If $\lim_{n\to\infty} Q_n = 0$, then

$$\lim_{n\to\infty} r(G, (\underline{\phi}_n^*, \underline{\gamma}^*_{n-1}, \underline{\delta}_n^*)) = r(G, (\underline{\phi}^*, \underline{\gamma}^*, \underline{\delta}^*)) \tag{2.71}$$

Proof: For any stopping time N, any integer j and any j-dimensional experiment rule $\underline{\gamma}^{j-1} = (\gamma_0, \gamma_1,\ldots, \gamma_{j-1})$, where each γ_i is $\mathcal{G}_i$-measurable and satisfies (2.13), define

$$T_j(N,\underline{\gamma}^{j-1}) = \int_{\{N=j\}} \{\rho_j[g_j(\underline{x}^j,\underline{e}^j)] + c_j[g_j(\underline{x}^j,\underline{e}^j)]\} f_{G,\underline{\gamma}^{j-1}}(\underline{x}^j,\underline{e}^j) \prod_{i=1}^{n} d\mu(e_i) \prod_{j=1}^{n} d\mu(x_j), \tag{2.72}$$

where

$$c_j[g_j(\underline{x}^j,\underline{e}^j)] = \int_\Theta c_j(\theta,\underline{x}^j,\underline{e}^j) g(\theta|\underline{x}^j,\underline{e}^j)\, d\nu(\theta) \tag{2.73}$$

(see (2.44)). Then it follows that

$$r(G, (\underline{\phi}^*,\underline{\gamma}^*,\underline{\delta}^*)) = \sum_{j=1}^{\infty} T_j(N^*,\underline{\gamma}^*_{j-1}) = \lim_{n\to\infty} \sum_{j=1}^{n} T_j(N^*,\underline{\gamma}^*_{j-1}) \tag{2.74}$$

Similarly, the risk for the truncated procedure $(\underline{\phi}^*_n, \underline{\gamma}^*_{n-1}, \underline{\delta}^*_n)$ may be written as

$$r(G, (\underline{\phi}^*_n, \underline{\gamma}^*_{n-1}, \underline{\delta}^*_n)) = \sum_{j=1}^{n} T_j(N^*_n, \underline{\gamma}^*_{j-1}) . \tag{2.75}$$

From the definition of N^*_n, however, it follows from (2.69) that $T_j(N^*_n, \underline{\gamma}^*_{j-1}) = T_j(N^*, \underline{\gamma}^*_{j-1})$ for $j = 1,2,\ldots, n-1$, and

$$T_n(N^*_n, \underline{\gamma}^*_{n-1}) = T_n(N^*, \underline{\gamma}^*_{n-1}) + Q_n + R_n \tag{2.76}$$

where Q_n is defined by (2.70) and

$$R_n = \int_{\{N^* > n\}} C_n[g_n(\underline{x}^n,\underline{e}^n)] f_{G,\underline{\gamma}^*_{n-1}}(\underline{x}^n,\underline{e}^n) \prod_{i=1}^{n} d\mu(e_i) \prod_{j=1}^{n} d\mu(x_j) . \tag{2.77}$$

Combining the above equations, it follows that

$$r(G, (\underline{\phi}^*_n, \underline{\gamma}^*_{n-1},\delta^*_n) = \sum_{j=1}^{n} T_j(N^*, \underline{\gamma}^*_{j-1}) + Q_n + R_n \tag{2.78}$$

Since $\underline{\phi}^*$ gives rise to a finite risk for all $\underline{\gamma}$ and $\underline{\gamma}^*$ in particular, however, it follows that $E_{f_\theta,\underline{\phi}^*,\underline{\gamma}^*}\{C_{N^*}[g_{N^*}(\underline{X}^{N^*},\underline{E}^{N^*})]\}$ is finite, and hence $R_n \to 0$ as $n \to \infty$. Therefore, if $Q_n \to 0$ as $n \to \infty$, the result follows from (2.74) and (2.78). □

Corollary 2.6.1. Either of the following statements is sufficient to

ensure that $Q_n \to 0$ as $n \to \infty$:

1. There exists a K such that, for all $(\underline{x}^n, \underline{e}^n) \in \{N^* > n\}$ and for all $n = 1,2,\ldots,$

$$\rho_n [g_n(\underline{x}^n, \underline{e}^n)] < nK \tag{2.79}$$

2. $$\lim_{n\to\infty} E_{G,f_\theta,\underline{\gamma}^*}\{\rho_n [g_n(\underline{X}^n, \underline{E}^n)]\} = 0. \tag{2.80}$$

Proof: If 1. is correct, then it follows from the definition of Q_n that

$$\lim_{n\to\infty} Q_n \leqslant \lim_{n\to\infty} nK\, P(N^* > n) = 0 \tag{2.81}$$

since $E_{G,f_\theta,\underline{\phi}^*,\underline{\gamma}^*}[N^*] < \infty$.

If 2. is correct, note that for each $n = 1,2,\ldots,$

$$E_{G,f_\theta,\underline{\gamma}^*}\{\rho_n [g_n(\underline{X}^n, \underline{E}^n)]\} = \left[\int_{\{N^* \leqslant n\}} + \int_{\{N^* > n\}}\right] \cdot \rho_n [g_n(\underline{x}^n, \underline{e}^n)]$$

$$\times f_{G,\underline{\gamma}^*_{n-1}}(\underline{x}^n, \underline{e}^n) \cdot \prod_{i=1}^{n} d\mu(e_i) \prod_{j=1}^{n} d\mu(x_j)$$

$$= P_n + Q_n, \quad \text{say} \tag{2.82}$$

Clearly P_n and Q_n are both non-negative, and so

$$Q_n \leqslant E_{G,f_\theta,\underline{\gamma}^*}\{\rho_n [g_n(\underline{X}^n, \underline{E}^n)]\} \tag{2.83}$$

and the result follows from (2.80). □

Corollary 2.6.2. If either condition 1. or condition 2. of Corollary 2.6.1. is satisfied, then

$$\lim_{n\to\infty} r(G, (\underline{\phi}^{n*}, \underline{\gamma}^{(n-1)*}, \underline{\delta}^{n*})) = r(G, (\underline{\phi}^*, \underline{\gamma}^*, \underline{\delta}^*)) \tag{2.84}$$

Proof: The truncated rules $\underline{\phi}^*_n$, $\underline{\gamma}^*_{n-1}$ and $\underline{\delta}^*_n$ are respectively, a stopping rule, an experiment rule, and a terminal decision rule for the problem

truncated after n observations. Thus, from the optimality property of $(\underline{\phi}^{n*}, \underline{\gamma}^{(n-1)*}, \underline{\delta}^{n*})$ stated in (2.61), and the optimality property of $(\underline{\phi}^{*}, \underline{\gamma}^{*}, \underline{\delta}^{*})$, it follows that

$$r(G, (\underline{\phi}^{*}, \underline{\gamma}^{*}, \underline{\delta}^{*})) \leqslant r(G, (\underline{\phi}^{n*}, \underline{\gamma}^{(n-1)*}, \underline{\delta}^{n*})) \leqslant r(G, (\phi_n^{*}, \gamma_{n-1}^{*}, \delta_n^{*}))$$
$$\text{for all } n = 1,2,\ldots . \qquad (2.85)$$

The result (2.84) then follows from Corollary 2.6.1 and Theorem 2.6. □

4. AN ESSENTIALLY COMPLETE CLASS OF PROCEDURES FOR THE SEQUENTIAL DESIGN PROBLEM

Although the existence of an optimal procedure for the non-truncated case has been shown, and the optimal procedure for the truncated case has been defined constructively, it is nevertheless of great benefit to the statistician if he can restrict the class of all sequential design procedures $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$, denoted as $\mathcal{P}$, to some subclass $\mathcal{P}^0$ which contains the optimal Bayes (or ε-Bayes) procedure. If the class $\mathcal{P}^0$ is easily identified in some sense, then the optimal procedure is more easily characterised, and hence more easily determined. The results of this section, which are based on the work of Gray (1968), show that $\mathcal{P}^0$ may be taken as the class of all procedures $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$ whose components are based on certain sufficient statistics of the arguments $(\underline{x}^n, \underline{e}^n)$ for each n, where the usual definition of sufficiency is appropriately modified.

Let $\mathcal{P}$ denote the class of all sequential design procedures $(\underline{\phi}^*, \underline{\gamma}^*, \underline{\delta}^*)$ for the problem based on the prior distribution G, where the components of $\underline{\gamma}^*$ are restricted to the class $\mathcal{E}^*$ defined previously. Given any two procedures $(\underline{\phi}_1, \underline{\gamma}_1, \underline{\delta}_1)$ and $(\underline{\phi}_2, \underline{\gamma}_2, \underline{\delta}_2)$, both in $\mathcal{P}$, $(\underline{\phi}_1, \underline{\gamma}_1, \underline{\delta}_1)$ is said to be <u>as good as</u> $(\underline{\phi}_2, \underline{\gamma}_2, \underline{\delta}_2)$ if

$$R(\theta, (\underline{\phi}_1, \underline{\gamma}_1, \underline{\delta}_1)) \leqslant R(\theta, (\underline{\phi}_2, \underline{\gamma}_2, \underline{\delta}_2)) \text{ for all } \theta \in \Theta \qquad (2.86)$$

where the risk function R was defined in (2.27). A class of sequential design procedures $\mathcal{P}^0$, $\mathcal{P}^0 \subset \mathcal{P}$, is said to be essentially complete if, given any procedure $(\underline{\phi}, \underline{\gamma}, \underline{\delta}) \in \mathcal{P}$ not in $\mathcal{P}^0$, there exists a rule $(\underline{\phi}^0, \underline{\gamma}^0, \underline{\delta}^0) \in \mathcal{P}$ which is as good as $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$.

It follows that if $(\underline{\phi}_1, \underline{\gamma}_1, \underline{\delta}_1)$ is as good as $(\underline{\phi}_2, \underline{\gamma}_2, \underline{\delta}_2)$, then

$$r(G, (\underline{\phi}_1, \underline{\gamma}_1, \underline{\delta}_1)) \leqslant r(G, (\underline{\phi}_2, \underline{\gamma}_2, \underline{\delta}_2)) \quad \text{for all prior distributions } G \tag{2.87}$$

and hence a Bayes or ε-Bayes procedure may be found by considering the class of procedures $\mathcal{P}^0$ only.

(1) Concepts of Sufficiency

A statistic T_j based on $(\underline{X}^j, \underline{E}^j)$ is defined as an $\mathcal{F}_j$-measurable function which maps $\mathcal{X}^j \times \mathcal{E}^j$ into a Euclidean space. A sequence of statistics $\{T_j\}$ where each T_j is based on $(\underline{X}^j, \underline{E}^j)$, is said to be parameter sufficient (PARS) if, for $j = 0,1,\ldots$, for all experiment rules $\underline{\gamma}$ and all sets $F_j \in \mathcal{F}_j$, there exists a version of $P_{f_\theta, \underline{\gamma}}(F_j | t_j)$ that is independent of θ. PARS may be characterised by the following version of the Neyman factorisation theorem:

Theorem 2.7. The sequence $\{T_j\}$ is PARS if and only if, for each $n = 1,2,\ldots$, there exists a non-negative $\mathcal{B}(T_n)$-measurable function $q_{\theta, \underline{\gamma}^{n-1}}(t_n)$ and a non-negative $\mathcal{F}_n$-measurable function $h_{\underline{\gamma}^{n-1}}(\underline{x}^n, \underline{e}^n)$ such that

$$f_{\theta, \underline{\gamma}^{n-1}}(\underline{x}^n, \underline{e}^n) = q_{\theta, \underline{\gamma}^{n-1}}[T_n(\underline{x}^n, \underline{e}^n)] h_{\underline{\gamma}^{n-1}}(\underline{x}^n, \underline{e}^n) \quad (\text{a.s. } \mathcal{F}_n) \tag{2.88}$$

Corollary 2.7.1. The sequence $\{T_j\}$ is PARS if and only if, for each $n = 1,2,\ldots$, there exists a non-negative $\mathcal{B}(T_n)$-measurable function $q_\theta(t_n)$ and a non-negative , $\mathcal{F}_n$-measurable function $h(\underline{x}^n, \underline{e}^n)$ such that

$$\prod_{j=1}^{n} f_\theta(x_j|\underline{x}^{j-1}, \underline{e}^j) = q_\theta(t_n)h(\underline{x}^n, \underline{e}^n) \quad \text{(a.s. } F_n\text{)}. \tag{2.89}$$

(Unless stated otherwise, the proofs of all results given in this section may be found in Gray (1968)).

The PARS of a sequence $\{T_j\}$ defined above is sufficient to show that $\{T_j\}$ is sequentially PARS, i.e. $\text{Dist}_{\theta,\underline{\phi},\underline{\gamma}}(X^N, \underline{E}^N|T_N, N)$ is independent of θ for all $\underline{\phi}$ and $\underline{\gamma}$. Formally,

Theorem 2.8. If $\{T_j\}$ is PARS, then for all $\underline{\phi}$, $\underline{\gamma}$ and all $F_n \in \mathcal{F}_n$, there exist versions of $P_{f_\theta,\underline{\phi},\underline{\gamma}}\{F_n|t_n, N=n\}$ and $P_{f_\theta,\underline{\phi},\underline{\gamma},\underline{\delta}}\{F_n|t_n, N \geqslant n\}$ that are independent of θ.

A further type of sufficiency is required before an essentially complete class of procedures can be defined; sufficiency of $\{T_j\}$ with respect to the rules or (as Gray calls them) policies $\underline{\phi}$ and $\underline{\gamma}$, as opposed to sufficiency with respect to the parameter θ. The sequence $\{T_j\}$ is said to be policy sufficient (POLS) if, for each $n = 0,1,\ldots$, for all θ and all $B_{n+1} \in \mathcal{B}(T_{n+1})$, there is a version of $P_{f_\theta,\underline{\phi},\underline{\gamma}}\{B_{n+1}|t_n, e_{n+1}, N \geqslant n+1\}$ that is independent of $\underline{\phi}$ and of $\underline{\gamma}$.

Theorem 2.9. If, for $n = 0,1,\ldots$, and for all $C_{n+1} \in \mathcal{B}(X_{n+1})$, there is a version of $P_{f_\theta,\underline{\phi},\underline{\gamma}}\{C_{n+1}|t_n, e_{n+1}, N \geqslant n+1\}$ which is independent of $\underline{\phi}$ and of $\underline{\gamma}$, for all $\theta \in \Theta$, and T_{n+1} is a $\mathcal{B}(T_n, X_{n+1}, E_{n+1})$-measurable function, then $\{T_j\}$ is POLS.

Corollary 2.9.1. If, for each $n = 0,1,\ldots$ and all $\theta \in \Theta$, T_{n+1} is $\mathcal{B}(T_n, X_{n+1}, E_{n+1})$-measurable and the conditional probability density $f_\theta(x_{n+1}|\underline{x}^n, \underline{e}^{n+1})$ is a $\mathcal{B}(T_n, X_{n+1}, E_{n+1})$-measurable function, say of the form $f_\theta(x_{n+1}|t_n, e_{n+1})$, then $\{T_j\}$ is POLS.

(2) Essential Completeness Of The Class Of Procedures Based On $\{T_j\}$

It may now be shown that the class of procedures $(\underline{\phi}^0, \underline{\gamma}^0, \underline{\delta}^0)$,

where each component rule is based on $\{T_j\}$, forms an essentially complete class of procedures for the sequential design problem, provided that $\{T_j\}$ is PARS and POLS.

A stopping rule $\underline{\phi}$ is said to be based on $\{T_j\}$ if, for each integer n,

$$\phi_n(\underline{x}^n, \underline{e}^n) = \phi_n(t_n) \tag{2.90}$$

is $\mathcal{B}(T_n)$-measurable. An experiment rule $\underline{\gamma}$ is said to be based on $\{T_j\}$ if, for each integer n,

$$\gamma_n(e_{n+1}|\underline{x}^n, \underline{e}^n) = \gamma_n(e_{n+1}|t_n) \tag{2.91}$$

is $\mathcal{B}(T_n, E_{n+1})$-measurable. A terminal decision rule $\underline{\delta}$ is said to be based on $\{T_j\}$ if, for each integer n,

$$\delta_n(a|\underline{x}^n, \underline{e}^n) = \delta_n(a|t_n) \tag{2.92}$$

is $\mathcal{B}(A, T_n)$-measurable.

Then Gray has shown

Theorem 2.10. If the sequence $\{T_j\}$ is PARS, then the class of procedures $\{\underline{\phi}, \underline{\gamma}, \underline{\delta}^0\}$, where $\underline{\delta}^0$ is based on $\{T_j\}$, is essentially complete.

Theorem 2.11. If $\{T_j\}$ is PARS and POLS and, for each integer n, $C_n(\theta, \underline{x}^n, \underline{e}^n) = C'_n(\theta, t_n)$ is $\mathcal{B}(T_n)$-measurable for all $\theta \in \Theta$, then the class of procedures $\{\underline{\phi}^0, \underline{\gamma}^0, \underline{\delta}^0\}$, based on $\{T_j\}$, is essentially complete.

CHAPTER III

APPLICATION OF BAYES SEQUENTIAL DESIGN THEORY TO AN IDENTIFICATION PROBLEM AND A RANKING PROBLEM

1. THE MULTIPLE DECISION PROBLEM

In this chapter the sequential design of experiments from a Bayesian viewpoint, and the results obtained in Chapter 2, are applied to the branch of statistical decision theory known as Multiple Decision Theory. In a multiple decision problem, the parameter θ is a vector $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ where each θ_i is a member of the same real subset $\Theta \subseteq R$, $i = 1,2,\ldots, k$, for some positive integer $k \geqslant 2$. Each θ_i may be thought of as being associated with a population Π_i from which successive independent r.v.s $X_1, X_2, \ldots$ may be observed according to a frequency function $f(x|\underline{\theta})$ whose known functional form is $f(x|\theta_i)$, $i = 1,2,\ldots, k$. The experiment space $\mathcal{E}$ is thus a finite space $\{e^1, e^2, \ldots, e^k\}$, where, for any stage $j = 1,2,\ldots$, $E_j = e^i$ means that the r.v. X observed at the j^{th} stage comes from population Π_i according to the frequency function $f(x|\theta_i)$, $i = 1,2,\ldots, k$. Furthermore, each population density $f_{\underline{\theta}}(x_{j+1}|\underline{x}^j, \underline{e}^{j+1})$ is of the form $f_{\underline{\theta}}(x_{j+1}|e_{j+1})$, namely $f(x_{j+1}|\theta_i)$ if $e_{j+1} = e^i$, as it is assumed that r.v.s from different populations are also independent. (Without loss of generality, $\mathcal{E}$ may be taken as $\{1,2,\ldots, k\}$).

The parameter space $\underline{\Theta} = \Theta^k$ may be thought of as partitioned into k disjoint subspaces Θ_i, $i = 1,2,\ldots, k$, where $\underline{\theta} \in \Theta_i$ iff the i^{th} component θ_i of $\underline{\theta}$ is greater than or equal to the remaining components; then the problem of deciding which subspace $\underline{\theta}$ lies in is equivalent to the problem of deciding which component θ_i is largest. Consequently, the action space A is also finite, $A = \{a_1, a_2, \ldots, a_k\}$, where a_i means "say $\underline{\theta} \in \Theta_i$", and the loss function $L(\underline{\theta}, a)$ satisfies

$$L(\underline{\theta}, a_i) = 0 \quad \text{if} \quad \underline{\theta} \in \Theta_i, \; i = 1,2,\ldots, k \tag{3.1}$$

with $L(\underline{\theta}, a_i) \geqslant 0$ otherwise.

Multiple decision problems may be classified into two types: identification problems and ranking problems. An identification problem is one in which the components of $\underline{\theta}$, $\{\theta_1, \theta_2,\ldots, \theta_k\}$, are known jointly, but the actual ordering is unknown; thus the parameter space Θ consists of the k! permutations* of the elements $\theta_1, \theta_2,\ldots, \theta_k$, and is finite. Solutions of identification problems are easily characterised by the symmetry that follows from the symmetry of the permutations; this will be seen in the first example considered. A ranking problem, however, is the more difficult case when $\underline{\Theta} = \Theta^k$ is infinite, Θ being typically an open interval on the real line. The symmetry of identification problems is also present in ranking problems, and so a solution to an identification problem may give insight into a corresponding solution for a ranking problem, or even provide a reasonable solution itself. Identification and ranking problems may be thought of as k-generalisations of tests of simple hypotheses and of composite hypotheses, respectively. Bechhofer, Kiefer, and Sobel (1968) contains a complete exposition of identification and ranking problems.

In this chapter the Bayes sequential design procedure is constructed for an identification problem and a ranking problem. The first, an identification problem concerning the parameters of several binomial populations, illustrates the savings in terms of risk, and in particular, expected sample size, which accrue if the optimal sequential design procedure is used instead of the usual optimal sequential decision procedure. The second, a ranking problem concerning the means of two normal populations, shows the difficulty of computing the optimal solution to a non-trivial problem, and hence indicates that an easily computed

*Provided the θ_i's are distinct, of course.

sub-optimal rule may be more desirable in most cases.

2. IDENTIFYING ONE OF SEVERAL HYPOTHESES CONCERNING THE PARAMETERS OF SEVERAL BINOMIAL POPULATIONS

(1) Set-up And Notation

The problem considered is that of 3 binomial populations; generalisation to the case $k \geqslant 3$ is straightforward. Denoting the i^{th} population as Π_i, and the associated parameter as θ_i, it follows that the parameter of interest is $\underline{\theta} = (\theta_1, \theta_2, \theta_3)$. It is assumed that $\underline{\Theta} = \{(\nu_2, \nu_1, \nu_1), (\nu_1, \nu_2, \nu_1), (\nu_1, \nu_1, \nu_2)\}$, where ν_1 and ν_2 $(\nu_2 > \nu_1)$ are two known values in the open interval (0,1). Writing $\underline{\Theta} = \{\underline{\theta}_1, \underline{\theta}_2, \underline{\theta}_3\}$ in the same ordering as above, it follows that $\underline{\theta} = \underline{\theta}_i$ iff $\theta_i = \nu_2$, $i = 1,2,3$, and the partition of $\underline{\Theta}$, $\{\Theta_1, \Theta_2, \Theta_3\}$, is defined by $\Theta_i = \{\underline{\theta}_i\}$, $i = 1,2,3$. The known prior probability distribution is a triple $\underline{g} = (g_1, g_2, g_3)$, where $g_i = P(\underline{\theta} = \underline{\theta}_i)$, $i = 1,2,3$, and $g_1 + g_2 + g_3 = 1$.

From each population Π_i it is possible to observe independent and identically distributed (i.i.d.) r.v.s $X_1, X_2, \ldots$, whose common probability distribution is given by

$$f(x|\theta_i) = \theta_i^x(1-\theta_i)^{1-x}, \quad x = 0,1 \quad , \quad i = 1,2,3 . \tag{3.2}$$

The action and experiment spaces are, respectively, $\{a_1, a_2, a_3\}$ and $\{1,2,3\}$, as defined in the previous section. The loss function is assumed to be zero-one, i.e.

$$L(\underline{\theta}, a_i) = \begin{cases} 0 & \text{if } \underline{\theta} \in \Theta_i \\ 1 & \text{otherwise} \end{cases} \quad , \; i = 1,2,3 \tag{3.3}$$

and the cost function $C_n(\underline{\theta}, \underline{x}^n, \underline{e}^n) = nc$ for all $\underline{\theta}$ and all $(\underline{x}^n, \underline{e}^n)$, where c is a positive constant. c is the (constant) cost per observation,

and the cost function so defined clearly satisfies (2.66) and all required measurability conditions.

(2) Sufficient Statistics

Before writing the expression for the Bayes conditional expected loss, it may be shown that a PARS and POLS sequence of statistics $\{T_j\}$ exists for this problem. Define the indicator function on $\mathcal{E}$

$$I_i(e) = \begin{cases} 1 & \text{if } e = i \\ 0 & \text{otherwise} \end{cases} \quad , \; i = 1,2,3 . \tag{3.4}$$

After n stages of experimentation, which yielded the values $(\underline{x}^n, \underline{e}^n)$, define the statistic $t_n = \{s_{1,n}, f_{1,n}, s_{2,n}, f_{2,n}, s_{3,n}, f_{3,n}\}$ by

$$\left.\begin{aligned} s_{i,n} &= \sum_{j=1}^{n} x_j I_i(e_j) \\ f_{i,n} &= \sum_{j=1}^{n} (1 - x_j) I_i(e_j) \end{aligned}\right\} \quad , \; i = 1,2,3 \tag{3.5}$$

$s_{i,n}$ (respectively, $f_{i,n}$) is the number of successes (respectively, failures) observed from population Π_i in the sequence $\{\underline{x}^n, \underline{e}^n\}$, $i = 1,2,3$, and is clearly $\mathcal{F}_n$-measurable. It follows that

$$\prod_{j=1}^{n} f_\theta(x_j|e_j) = \theta_1^{s_{1,n}}(1-\theta_1)^{f_{1,n}} \theta_2^{s_{2,n}}(1-\theta_2)^{f_{2,n}} \theta_3^{s_{3,n}}(1-\theta_3)^{f_{3,n}} \tag{3.6}$$

and hence $\{t_j\}$ is PARS by Corollary 2.7.1. Furthermore, writing

$$\left.\begin{aligned} s_{i,n+1} &= s_{i,n} + x_{n+1} I_i(e_{n+1}) \\ f_{i,n+1} &= f_{i,n} + (1 - x_{n+1}) I_i(e_{n+1}) \end{aligned}\right\} \quad , \; i = 1,2,3 \tag{3.7}$$

it is seen that $\{t_j\}$ is POLS by Corollary 2.9.1. It follows that the posterior density $g(\underline{\theta}_1 \mid \underline{x}^n, \underline{e}^n)$, say, may be written as

$$g(\underline{\theta}_1|t_n) = \frac{g_1\nu_2^{s_{1,n}}(1-\nu_2)^{f_{1,n}}\nu_1^{s_{2,n}+s_{3,n}}(1-\nu_1)^{f_{2,n}+f_{3,n}}}{\left[g_1\nu_2^{s_{1,n}}(1-\nu_2)^{f_{1,n}}\nu_1^{s_{2,n}+s_{3,n}}(1-\nu_1)^{f_{2,n}+f_{3,n}}\right.}$$

$$+ g_2\nu_2^{s_{2,n}}(1-\nu_2)^{f_{2,n}}\nu_1^{s_{1,n}+s_{3,n}}(1-\nu_1)^{f_{1,n}+f_{3,n}}$$

$$\left. + g_3\nu_2^{s_{3,n}}(1-\nu_2)^{f_{3,n}}\nu_1^{s_{1,n}+s_{2,n}}(1-\nu_1)^{f_{1,n}+f_{2,n}}\right]$$

$$= \frac{F_{1,n}}{F_{1,n} + F_{2,n} + F_{3,n}} \quad , \quad \text{say} \tag{3.8}$$

for notational convenience. Then the posterior probability of $\underline{\theta}_i$ given n observations is

$$g(\underline{\theta}_i|t_n) = \frac{F_{i,n}}{\sum_{j=1}^{3} F_{j,n}} \quad , \quad i = 1,2,3 . \tag{3.9}$$

From the symmetry of the arguments $\underline{\theta}_i$, it is sufficient to consider one argument only. Thus the expected posterior loss after n observations if a_1 is chosen is, from (2.41)

$$\int_{\underline{\Theta}} L(\underline{\theta}, a_1)\, g(\underline{\theta}|t_n)\, d\nu(\underline{\theta})$$

$$= \left[\int_{\Theta_2} + \int_{\Theta_3}\right] g(\underline{\theta}|t_n)\, d\nu(\underline{\theta})$$

by definition of the 0 - 1 loss function (3.3),

$$= g(\underline{\theta}_2|t_n) + g(\underline{\theta}_3|t_n)$$

$$= \frac{F_{2,n} + F_{3,n}}{\sum_{j=1}^{3} F_{j,n}} \tag{3.10}$$

Thus the minimum expected posterior loss plus cost after n observations is

$$\frac{\sum_{j=1}^{3} F_{j,n} - \max_i F_{i,n}}{\sum_{j=1}^{3} F_{j,n}} + nc \tag{3.11}$$

In this case, the computation of the optimal sequential design procedure and the Bayes risk for the truncated problem, using backward induction, is reasonably straightforward. Firstly, since $\mathcal{E}$ is finite, the "M" functions defined by (2.61) may be evaluated by computing the appropriate expectation for each $e \in \mathcal{E}$ and taking the minimum. Secondly, the density $f_G(x_j | \underline{x}^{j-1}, \underline{e}^j)$ required for this expectation is a 2-point distribution, as x_j can only be 0 or 1. Therefore, providing the truncation number J is not too large, it is possible, at each stage $j = 0,1,\ldots, J-1$, to compute the values of $U_j(g_j)$ and $M_{V_{j+1}(\cdot)}$ for each possible value of t_j. The results of these calculations for various values of ν_1, ν_2, J, and c are given in Table 3.1 below, which tabulates values of the Bayes risk and the expected sample size. (The quantity R is used to weight the expected loss against a unit cost c = 1, and may be thought of as 1/c: thus $U_j = R\rho_j + j$ instead of $\rho_j + jc$). Also included are the corresponding values for the sequential decision problem without a choice of experiments; at each stage j the statistician may stop and receive a terminal loss plus cost or observe the vector $X_j = (X_{1,j}, X_{2,j}, X_{3,j})$ where $X_{i,j}$ is observed from population Π_i according to $f(x|\theta_i)$, i = 1,2,3. Therefore, the statistician either stops or observes 3 observations, one from each population, at a cost of 3c units. Consequently, the number of "observations" taken is a multiple of 3, and so J is chosen as a multiple of 3 in this case.

TABLE 3.1 Values of the Bayes risk V_0 and the expected sample size $E[N]$ for the Bayes sequential design and decision rules. (3 populations) ($g = (\cdot 5, \cdot 3, \cdot 2)$)

J	R	ν_1	ν_2	Design		Non-design		Percentage Reduction	
				$V_0(\underline{\gamma}^{(J-1)*})$	$E[N]$	$V_0^{(J-1)}$	$E[N]$	Risk	$E[N]$
5	20	.2	.7	8.2217	3.372	-	-		
4	20	.2	.7	8.5697	3.110	-	-		
3	20	.2	.7	7.7820	1.870	8.9280	3.000	12.84	37.67
3	10	.2	.7	4.8260	1.870	5.0000	3.000	3.48	37.67
3	50	.2	.7	16.3500	2.385	17.8200	3.000	8.25	20.50
3	100	.2	.7	30.0500	2.670	32.6400	3.000	7.94	11.00
3	50	.3	.7	19.8100	1.766	21.0450	3.000	5.87	41.13
3	200	.4	.6	95.1200	1.680	97.4000	3.000	2.34	44.00
3	50	.1	.6	15.1350	2.800	16.9400	3.000	10.66	6.67
3	50	.1	.3	23.5320	2.152	24.8450	3.000	5.28	28.27

Because the experimentation process in the decision problem uses a fixed design (viz. observe the next observation from the next population), it follows that the Bayes risk for the sequential design problem is always no larger than the Bayes risk for the corresponding sequential decision problem, as is seen in the table; the percentage reduction in the risk is also shown. However, comparison of the expected sample sizes reveals a much larger saving; this implies that the reduction in risk by using an optimal sequential design is mostly due to the reduction in sampling cost. This is hardly surprising, however, as the decision rule is restricted to either stopping without taking any observations or observing 3 observations (in general, k observations), and will always do the latter provided that R is sufficiently large. Unfortunately, computations for J = 6, 9,...

were not possible because of the storage space required on the computer, but in Table 3.2 results have been tabulated for the case of 2 populations,

TABLE 3.2 Values of the Bayes risk V_0 and the expected sample size $E[N]$ for the Bayes sequential design and decision rules (2 populations) ($\underline{g} = (.7, .3)$)

J	R	ν_1	ν_2	Design		Non-design		Percentage Reduction	
				$V_0(\underline{\gamma}^{(J-1)*})$	$E[N]$	$V_0^{(J-1)}$	$E[N]$	Risk	$E[N]$
4	50	.2	.7	8.4489	2.933	8.8140	3.180	4.14	7.18
6	50	.2	.7	7.6019	3.467	7.8571	3.855	3.25	10.07
8	50	.2	.7	7.2118	3.753	7.3388	4.238	1.73	11.43
6	200	.4	.6	55.2409	3.457	55.6061	3.822	0.65	9.55
6		.1	.6	6.8265	2.783	7.2997	3.935	6.48	29.27
6		.1	.3	25.4462	3.809	26.312	4.832	3.29	21.18

($\underline{\Theta} = \{(\nu_2, \nu_1), (\nu_1, \nu_2)\}$), in order that the performance of both rules may be compared for J = 4,6,8. From these results it is observed that, although the percentage reduction in risk from using the Bayes design rule decreases as J increases, the percentage reduction in expected sample size increases. Moreover, this reduction will increase with k, the number of populations. Therefore, it may be concluded that for comparable values of the risk, the Bayes sequential design procedure has the potential to reduce the average number of observations required (i.e. the expected sampling cost) by a large amount. In many situations, a certain level of risk is considered acceptable and the goal is to attain that level using as few observations as is necessary; in such situations, the Bayes sequential design procedure will often be preferable to the Bayes sequential decision procedure, despite the extra computation involved.

The FORTRAN program used to generate $V_0(\gamma^{(J-1)*})$ and $E[N]$ for the 2-population case is given in Appendix A2.

3. SELECTING FROM TWO NORMAL POPULATIONS THE POPULATION WITH THE LARGER MEAN UNDER LINEAR LOSS; A RANKING PROBLEM

(1) Set-up And Notation

Consider two populations Π_1 and Π_2 with associated parameters θ_1 and θ_2, respectively. It is assumed that Θ is the real line, i.e., $\underline{\Theta} = \{(\theta_1, \theta_2) : -\infty < \theta_1, \theta_2 < \infty\}$, and hence the partition $\{\Theta_1, \Theta_2\}$ is defined by $\Theta_1 = \{(\theta_1, \theta_2) \in \underline{\Theta} : \theta_1 \geqslant \theta_2\}$ and $\Theta_2 = \{(\theta_1, \theta_2) \in \underline{\Theta} : \theta_1 \leqslant \theta_2\}$, where Θ_1 and Θ_2 have the boundary $\theta_1 = \theta_2$ in common. Further, it is also assumed that each θ_i possesses a Normal prior distribution with mean λ_i and variance β_i^2 both known (henceforth written as $N(\lambda_i, \beta_i^2)$), $i = 1,2$, and that θ_1 and θ_2 are independent.

Both populations are Normal; that is, it is possible to observe i.i.d. r.v.'s $X_1, X_2, \ldots$ from Π_1 with common distribution $N(\theta_1, \sigma_1^2)$, and it is possible to observe i.i.d. r.v.'s $Y_1, Y_2, \ldots$ from Π_2 with common distribution $N(\theta_2, \sigma_2^2)$, where σ_1^2 and σ_2^2 are both known; in the usual notation

$$f(x|\theta_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left\{-\frac{(x-\theta_1)^2}{2\sigma_1^2}\right\}, \quad x \in R$$

$$f(y|\theta_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2} \exp\left\{-\frac{(y-\theta_2)^2}{2\sigma_2^2}\right\}, \quad y \in R. \tag{3.12}$$

It is also assumed that X_i (given θ_1) is independent of Y_j (given θ_2), $i,j = 1,2,\ldots$. The action and experiment spaces are thus $\{a_1, a_2\}$ and $\{1,2\}$ respectively, and the <u>linear loss</u> function is defined by

$$L(\underline{\theta}, a_1) = \begin{cases} K(\theta_2 - \theta_1) & \text{if } \underline{\theta} \in \Theta_2 \\ 0 & \text{if } \underline{\theta} \in \Theta_1 \end{cases},$$

$$L(\underline{\theta}, a_2) = \begin{cases} K(\theta_1 - \theta_2) & \text{if} \quad \underline{\theta} \in \Theta_1 \\ 0 & \text{if} \quad \underline{\theta} \in \Theta_2 \end{cases} \tag{3.13}$$

where K is a known positive constant whose value is assumed to be 1 without loss of generality. Finally, a constant cost per observation c is assumed.

(2) Sufficient Statistics

Once again, a PARS and POLS sequence of statistics $\{T_j\}$ may be found for this problem. Using the indicator function (3.4), the joint density function may be written as

$$\begin{aligned} \prod_{j=1}^{n} f_{\underline{\theta}}(\hat{x}_j | e_j) = & \left[\left(\frac{1}{\sqrt{2\pi}\ \sigma_1} \right) \exp\left\{ - \frac{\theta_1^2}{2\sigma_1^2} \right\} \right]^{\sum_{j=1}^{n} I_1(e_j)} \\ & \times \left[\left(\frac{1}{\sqrt{2\pi}\ \sigma_2} \right) \exp\left\{ - \frac{\theta_2^2}{2\sigma_2^2} \right\} \right]^{\sum_{j=1}^{n} I_2(e_j)} \\ & \times \exp\left\{ \frac{\theta_1}{\sigma_1^2} \sum_{j=1}^{n} \hat{x}_j I_1(e_j) + \frac{\theta_2}{\sigma_2^2} \sum_{j=1}^{n} \hat{x}_j I_2(e_j) \right\} \\ & \times \exp\left\{ - \frac{1}{2\sigma_1^2} \sum_{j=1}^{n} \hat{x}_j^2 I_1(e_j) - \frac{1}{2\sigma_2^2} \sum_{j=1}^{n} \hat{x}_j^2 I_2(e_j) \right\} \end{aligned} \tag{3.14}$$

for n = 1,2,... . Denoting $n_i = \sum_{j=1}^{n} I_i(e_j)$, i = 1,2, it follows that the sequence of statistics $\{T_j\}$ defined by

$$\begin{aligned} T_n &= \left(n_1, n_2, \sum_{i=1}^{n_1} X_i, \sum_{j=1}^{n_2} Y_j \right) \\ T_{n+1} &= T_n + (I_1(E_{n+1}), I_2(E_{n+1}), \hat{X}_{n+1} I_1(E_{n+1}), \hat{X}_{n+1} I_2(E_{n+1})), \\ & \qquad n = 1,2,\ldots , \end{aligned} \tag{3.15}$$

is PARS and POLS for the sequential design problem, where $(X_1,\ldots, X_{n_1}, Y_1,\ldots, Y_{n_2})$ is the permutation of the vector $(\hat{X}_1,\ldots, \hat{X}_n)$

of observation r.v.s corresponding to the two populations Π_1 and Π_2 from which $\{X_i\}$ and, respectively, $\{Y_j\}$, are observed. This result corresponds to the usual result concerning sufficient statistics for a sequence of i.i.d. r.v.s from a joint Normal density.

However, it is more convenient to use the sufficient statistic $T_n' = (\underset{n_1}{\underline{X}}, \underset{n_2}{\underline{Y}}) = (X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2})$, and henceforth m and n will be used to denote the number of observations from Π_1, and Π_2, respectively, out of a total of m+n observations, i.e. the problem is considered in terms of the individual population sample sizes, rather than the overall sample size. Thus $(\underline{X}^m, \underline{Y}^n)$ will be referred to as (m,n) r.v.s.

Given $(\underline{X}^m, \underline{Y}^n) = (\underline{x}^m, \underline{y}^n)$, it follows from the independence assumed above that

$$f(\underline{x}^m, \underline{y}^n|\underline{\theta}) = \prod_{i=1}^{m} f(x_i|\theta_1) \cdot \prod_{j=1}^{n} f(y_j|\theta_2) = f(\underline{x}^m|\theta_1) \cdot f(\underline{y}^n|\theta_2), \quad (3.16)$$

and also the posterior density $g(\underline{\theta}|\underline{x}^m, \underline{y}^n)$ is the product of the two densities $g_1(\theta_1|\underline{x}^m)$ and $g_2(\theta_2|\underline{y}^n)$, where

$$g_1(\theta_1|\underline{x}^m) = \frac{g_1(\theta_1) \prod_{i=1}^{m} f(x_i|\theta_1)}{\int_\Theta g_1(\theta) \prod_{i=1}^{m} f(x_i|\theta) d\theta} \quad * \quad (3.17)$$

and a similar expression holds for $g_2(\theta_2|\underline{y}^n)$. $(g_1|\theta_1)$ is the prior density $N(\lambda_1, \beta_1^2))$. Therefore, θ_1 and θ_2 are independent a posteriori as well as a priori. Furthermore, it is well-known (see e.g. Lindley (1965), p.3) that $g(\theta_1|\underline{x}^m)$ and $g(\theta_2|\underline{y}^n)$ are respectively, $N(\rho_{1m}, \tau_{1m}^2)$ and $N(\rho_{2n}, \tau_{2n}^2)$, where

$$\rho_{1m} = \frac{\beta_1^2 \sum_{i=1}^{m} x_i + \sigma_1^2 \lambda_1}{m\beta_1^2 + \sigma_1^2} \quad , \quad \rho_{2n} = \frac{\beta_2^2 \sum_{j=1}^{n} y_j + \sigma_2^2 \lambda_2}{n\beta_2^2 + \sigma_2^2}$$

$$\tau_{1m}^2 = \frac{\sigma_1^2 \beta_1^2}{m\beta_1^2 + \sigma_1^2} \quad , \quad \tau_{2n}^2 = \frac{\sigma_2^2 \beta_2^2}{n\beta_2^2 + \sigma_2^2} \quad (3.18)$$

$\rho_{i.}$ and $\tau_{i.}^2$ are the posterior mean and posterior variance of θ_i, i = 1,2.

*with respect to Lebesgue measure on the real line

(2) Evaluation Of Expected Loss

From the linear loss function (3.13), it follows that the expected loss after (m,n) observations if a_1 (resp., a_2) is chosen, denoted by $L^{(1)}_{m,n}$ (resp., $L^{(2)}_{m,n}$), may be written as

$$\left.\begin{aligned} L^{(1)}_{m,n} &= \int_{\Theta_2} (\theta_2 - \theta_1)\, g(\underline{\theta} \mid \underline{x}^m, \underline{y}^n)\, d\underline{\theta} \\ L^{(2)}_{m,n} &= \int_{\Theta_1} (\theta_1 - \theta_2)\, g(\underline{\theta} \mid \underline{x}^m, \underline{y}^n)\, d\underline{\theta} \end{aligned}\right\} \tag{3.19}$$

Note that

$$\begin{aligned} L^{(1)}_{m,n} - L^{(2)}_{m,n} &= \int_{\underline{\Theta}} (\theta_2 - \theta_1)\, g(\underline{\theta} \mid \underline{x}^m, \underline{y}^n)\, d\underline{\theta} \\ &= \int_{\Theta} \theta_2 g_2(\theta_2 \mid \underline{y}^n)\, d\theta_2 - \int_{\Theta} \theta_1 g_1(\theta_1 \mid \underline{x}^m)\, d\theta_1 \\ &\qquad \text{since } \theta_1 \text{ and } \theta_2 \text{ are independent,} \\ &= \rho_{2n} - \rho_{1m}. \end{aligned} \tag{3.20}$$

Thus the Bayes terminal decision rule after (m,n) observations, which is to choose the action associated with the smaller expected loss, is equivalent to choosing the action associated with the larger posterior mean; that is, θ_1 is said to be greater than or equal to θ_2 if $E[\theta_1 \mid \underline{x}^m] \geqslant E[\theta_2 \mid \underline{y}^n]$ and vice versa, a result which agrees with intuition, and is also true for the corresponding decision problem (see Deely (1965), Theorem 3.14). Consequently, it is necessary to evaluate only one of the $L^{(i)}_{m,n}$'s, the other following from (3.20). Considering $L^{(1)}_{m,n}$, write

$$L^{(1)}_{m,n} = \int_{\{(\theta_1,\theta_2) \,:\, \theta_1 \leqslant \theta_2\}} (\theta_2 - \theta_1)\, g_1(\theta_1 \mid \underline{x}^m)\, g_2(\theta_2 \mid \underline{y}^n)\, d\theta_1 d\theta_2 \tag{3.21}$$

This may be solved by a change of variables. Letting

$$\gamma_1 = \frac{\theta_2 - \theta_1}{\sqrt{2}} \quad , \quad \gamma_2 = \frac{\theta_2 + \theta_1}{\sqrt{2}} \tag{3.22}$$

it follows that γ_1 is $N((\rho_{2n} - \rho_{1m})/\sqrt{2},\ (\tau_{1m}^2 + \tau_{2n}^2)/2)$ and γ_2 is $N((\rho_{2n} + \rho_{1m})/\sqrt{2},\ (\tau_{1m}^2 + \tau_{2n}^2)/2)$. The inverse transformations are

$$\theta_1 = \frac{\gamma_2 - \gamma_1}{\sqrt{2}} = H_1(\gamma_1, \gamma_2), \quad \theta_2 = \frac{\gamma_2 + \gamma_1}{\sqrt{2}} = H_2(\gamma_1, \gamma_2), \text{ say} \tag{3.23}$$

and hence the Jacobian of the transformation is

$$J(\gamma_1, \gamma_2) = \begin{vmatrix} \frac{\partial H_1}{\partial \gamma_1} & \frac{\partial H_1}{\partial \gamma_2} \\ \frac{\partial H_2}{\partial \gamma_1} & \frac{\partial H_2}{\partial \gamma_2} \end{vmatrix} = -\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}$$

$$= -1, \quad \text{i.e.} \quad |J| = 1. \tag{3.24}$$

Therefore

$$\begin{aligned} L_{m,n}^{(1)} &= \int_{-\infty}^{\infty} \left[\int_0^{\infty} \sqrt{2}\ \gamma_1\ h(\gamma_1, \gamma_2 | \underline{x}^m, \underline{y}^n)\, d\gamma_1 \right] d\gamma_2 \\ &= \int_0^{\infty} \sqrt{2}\ \gamma_1 \left[\int_{-\infty}^{\infty} h(\gamma_1, \gamma_2 | \underline{x}^m, \underline{y}^n)\, d\gamma_2 \right] d\gamma_1 \\ &= \int_0^{\infty} \sqrt{2}\ \gamma_1\ h_1(\gamma_1 | \underline{x}^m, \underline{y}^n)\, d\gamma_1 \end{aligned} \tag{3.25}$$

where h is the joint density function of γ_1 and γ_2, and h_1 is the (Normal) density of γ_1. Defining

$$\eta_{m,n} = \rho_{2n} - \rho_{1m} \quad , \quad \xi_{m,n}^2 = \tau_{1m}^2 + \tau_{2n}^2 \tag{3.26}$$

and making the change of variable $t = (\sqrt{2}\ \gamma_1 - \eta_{m,n})/\xi_{m,n}$, it follows from (3.25) that

$$\begin{aligned} L_{m,n}^{(1)} &= \int_{-\eta_{m,n}/\xi_{m,n}}^{\infty} (\xi_{m,n} t + \eta_{m,n})\, \phi(t)\, dt \\ &= \xi_{m,n}\ \phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) + \eta_{m,n}\ \Phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) \end{aligned} \tag{3.27}$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{and} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt \tag{3.28}$$

denote the density function and distribution function, respectively, of a standard Normal (N(0,1)) r.v. From (3.20), (3.26) and (3.27) it follows that

$$L^{(2)}_{m,n} = \xi_{m,n}\,\phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) + \eta_{m,n}\,\Phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) - \eta_{m,n}$$

$$= \xi_{m,n}\,\phi\left(\frac{-\eta_{m,n}}{\xi_{m,n}}\right) - \eta_{m,n}\,\Phi\left(\frac{-\eta_{m,n}}{\xi_{m,n}}\right) \qquad (3.29)$$

Therefore, if the real-valued function $Q(x,y)$ is defined by

$$Q(x,y) = \begin{cases} y\,\phi\left(\frac{x}{y}\right) + x\,\Phi\left(\frac{x}{y}\right) & \text{if } x \leqslant 0 \\ y\,\phi\left(\frac{-x}{y}\right) - x\,\Phi\left(\frac{-x}{y}\right) & \text{if } x \geqslant 0 \end{cases} \qquad (3.30)$$

where $x \in R$, $y \in R^+$, then, with an obvious change in notation, the minimum expected loss plus cost after (m,n) observations $(\underline{x}^m, \underline{y}^n)$ is

$$U_{m,n}(g_{m,n}) = Q(\eta_{m,n}, \xi_{m,n}) + (m+n)c \qquad (3.31)$$

(3) Evaluation Of Expected Risk From Taking Another Observation

Having taken (m,n) observations and having obtained an expression for the loss plus cost $U_{m,n}$ in (3.31) above, it is now necessary to obtain an expression for the expected loss plus cost if a further observation, say X_{m+1}, is taken, and then a terminal decision is made. Define

$$\rho_{1(m+1)}(X_{m+1}) = \frac{\beta_1^2\left(\sum_{i=1}^{m} x_i + X_{m+1}\right) + \sigma_1^2\,\lambda_1}{(m+1)\beta_1^2 + \sigma_1^2},$$

$$\eta_{m+1,n}(X_{m+1}) = \rho_{2n} - \rho_{1(m+1)}(X_{m+1}) \qquad (3.32)$$

and

$$L^{(1)}_{m+1,n}(X_{m+1}) = \xi_{m+1,n}\,\phi\left(\frac{\eta_{m+1,n}(X_{m+1})}{\xi_{m+1,n}}\right) + \eta_{m+1,n}(X_{m+1})\,\Phi\left(\frac{\eta_{m+1,n}(X_{m+1})}{\xi_{m+1,n}}\right)$$

(3.33)

with a similar expression for $L^{(2)}_{m+1,n}(X_{m+1})$. It follows from (3.20) that

$$L^{(1)}_{m+1,n}(x) < L^{(2)}_{m+1,n}(x)$$

iff

$$\rho_{1(m+1)}(x) > \rho_{2n}$$

iff

$$x > \frac{\sigma_1^2(\rho_{2n}\tau^2_{1m} - \rho_{1m}\tau^2_{1(m+1)})}{\tau^2_{1m}\tau^2_{1(m+1)}} = k^{(1)}_{m+1,n} \quad , \text{ say} \tag{3.34}$$

where the last step follows from the identity

$$\rho_{1(m+1)}(x) = \frac{\rho_{1m}\tau^2_{1(m+1)}}{\tau^2_{1m}} + \frac{x\tau^2_{1(m+1)}}{\sigma_1^2} \, . \tag{3.35}$$

Therefore, the expected loss plus cost of observing X_{m+1} and then stopping is

$$E\left\{U_{m+1,n}\left[g_{m,n}(X_{m+1})\right] \middle| \underline{x}^m, \underline{y}^n\right\} = E\left[\min_{i=1,2}\left\{L^{(i)}_{m+1,n}(X_{m+1})\right\} \middle| \underline{x}^m, \underline{y}^n\right] + (m+n+1)c, \tag{3.36}$$

where

$$\begin{aligned} &E\left[\min_{i=1,2}\left\{L^{(i)}_{m+1,n}(X_{m+1})\right\} \middle| \underline{x}^m, \underline{y}^n\right] \\ &= \int_{-\infty}^{k^{(1)}_{m+1,n}} L^{(2)}_{m+1,n}(x) f_G(x|\underline{x}^m, \underline{y}^n)\,dx \\ &\quad + \int_{k^{(1)}_{m+1,n}}^{\infty} L^{(1)}_{m+1,n}(x) . f_G(x|\underline{x}^m, \underline{y}^n)\,dx \\ &= \int_{-\infty}^{\infty} L^{(1)}_{m+1,n}(x) . f_G(\underline{x}|\underline{x}^m, \underline{y}^n)\,dx - \int_{-\infty}^{k^{(1)}_{m+1,n}} \eta_{m+1,n}(x) f_G(x|\underline{x}^m, \underline{y}^n)\,dx \end{aligned} \tag{3.37}$$

where $f_G(x|\underline{x}^m, \underline{y}^n)$ is the density function of x_{m+1} conditional on $(\underline{x}^m, \underline{y}^n)$ but unconditional on $\underline{\theta}$ (see (2.45) ff.). It may easily be shown that $f_G(x|\underline{x}^m, \underline{y}^n)$ is of the form $f_G(x|\underline{x}^m)$, being the Normal density $N(\rho_{1m}, \tau^2_{1m} + \sigma_1^2)$.

Therefore, there are three integrals to evaluate.

(a) Evaluation Of The Integral

$$\xi_{m+1,n} \int_{-\infty}^{\infty} \phi\left(\frac{\eta_{m+1,n}(x)}{\xi_{m+1,n}}\right) f_G(x|\underline{x}^m)\,dx \tag{3.38}$$

Ignoring all "non-exponential" constants, it follows that -2 log(integrand) equals

$$\begin{aligned}
&\frac{\eta^2_{m+1,n}(x)}{\xi^2_{m+1,n}} + \frac{(x-\rho_{1m})^2}{\tau^2_{1m}+\sigma^2_1} \\
&= \frac{\rho^2_{2n}}{\tau^2_{1(m+1)}+\tau^2_{2n}} - \frac{2\rho_{2n}\tau^2_{1(m+1)}}{\sigma^2_1(\tau^2_{1(m+1)}+\tau^2_{2n})}\,x - \frac{2\rho_{1m}\rho_{2n}\tau^2_{1(m+1)}}{\tau^2_{1m}(\tau^2_{1(m+1)}+\tau^2_{2n})} \\
&\quad + \frac{\tau^4_{1(m+1)}}{\sigma^4_1(\tau^2_{1(m+1)}+\tau^2_{2n})}\,x^2 + \frac{2\tau^4_{1(m+1)}\rho_{1m}}{\tau^2_{1m}\sigma^2_1(\tau^2_{1(m+1)}+\tau^2_{2n})}\,x + \frac{\tau^4_{1(m+1)}\rho^2_{1m}}{\tau^4_{1m}(\tau^2_{1(m+1)}+\tau^2_{2n})} \\
&\quad + \frac{x^2}{\tau^2_{1m}+\sigma^2_1} - \frac{2\rho_{1m}}{\tau^2_{1m}+\sigma^2_1}\,x + \frac{\rho^2_{1m}}{\tau^2_{1m}+\sigma^2_1}
\end{aligned} \tag{3.39}$$

using (3.32) and (3.35). Therefore

$$\begin{aligned}
\text{coefficient of } x^2 &= \frac{1}{\tau^2_{1m}+\sigma^2_1} + \frac{\tau^4_{1(m+1)}}{\sigma^4_1(\tau^2_{1(m+1)}+\tau^2_{2n})} \\
&= \frac{\sigma^4_1(\tau^2_{1(m+1)}+\tau^2_{2n}) + \tau^4_{1(m+1)}(\tau^2_{1m}+\sigma^2_1)}{\sigma^4_1(\tau^2_{1m}+\sigma^2_1)(\tau^2_{1(m+1)}+\tau^2_{2n})} \\
&= B, \quad \text{say}
\end{aligned}$$

$$\begin{aligned}
\text{coefficient of } x &= \frac{-2\rho_{2n}\tau^2_{1(m+1)}}{\sigma^2_1(\tau^2_{1(m+1)}+\tau^2_{2n})} + \frac{2\tau^4_{1(m+1)}\rho_{1m}}{\tau^2_{1m}\sigma^2_1(\tau^2_{1(m+1)}+\tau^2_{2n})} - \frac{2\rho_{1m}}{\tau^2_{1m}+\sigma^2_1} \\
&= -2A, \quad \text{say}
\end{aligned}$$

$$\text{constants} = \frac{\rho^2_{2n}}{\tau^2_{1(m+1)}+\tau^2_{2n}} - \frac{2\rho_{1m}\rho_{2n}\tau^2_{1(m+1)}}{\tau^2_{1m}(\tau^2_{1(m+1)}+\tau^2_{2n})}$$

$$+ \frac{\tau_{1(m+1)}^4 \rho_{1m}^2}{\tau_{1m}^4 (\tau_{1(m+1)}^2 + \tau_{2n}^2)} + \frac{\rho_{1m}^2}{\tau_{1m}^2 + \sigma_1^2}$$

$$= C, \quad \text{say} \tag{3.40}$$

Therefore, completing the square, -2 log(integrand) equals

$$B\left(x - \frac{A}{B}\right)^2 + C - \frac{A^2}{B} \tag{3.41}$$

and hence integral (3.38) equals

$$\frac{\sqrt{\tau_{1(m+1)}^2 + \tau_{2n}^2}}{\sqrt{2\pi}\sqrt{\tau_{1m}^2 + \sigma_1^2}\sqrt{B}} \exp\left\{-\frac{1}{2}\frac{CB - A^2}{B}\right\} . \tag{3.42}$$

Direct calculation yields (after many tedious steps)

$$C - \frac{A^2}{B} = \frac{\left[\rho_{1m}\tau_{1(m+1)}^2(\tau_{1m}^2 + \sigma_1^2) - \rho_{2n}\tau_{1m}^2\sigma_1^2\right]^2}{\tau_{1m}^4\left[\sigma_1^4(\tau_{1(m+1)}^2 + \tau_{2n}^2) + \tau_{1(m+1)}^4(\tau_{1m}^2 + \sigma_1^2)\right]} . \tag{3.43}$$

From the definition (3.18) of τ_{1m}^2, however, it follows that the identity

$$\tau_{1(m+1)}^2(\tau_{1m}^2 + \sigma_1^2) = \tau_{1m}^2\sigma_1^2 \tag{3.44}$$

holds. Therefore, the numerator in (3.43) equals

$$\tau_{1m}^4\sigma_1^4\eta_{m,n}^2 \tag{3.45}$$

and the denominator equals

$$\tau_{1m}^4\sigma_1^2\left[\sigma_1^2(\tau_{1(m+1)}^2 + \tau_{2n}^2) + \tau_{1(m+1)}^2\tau_{1m}^2\right]$$

$$= \tau_{1m}^4\sigma_1^2\left[\tau_{2n}^2\sigma_1^2 + \tau_{1(m+1)}^2(\tau_{1m}^2 + \sigma_1^2)\right]$$

$$= \tau_{1m}^4\sigma_1^4(\tau_{1m}^2 + \tau_{2n}^2), \tag{3.46}$$

using (3.44) again. Therefore

$$C - \frac{A^2}{B} = \frac{\eta_{m,n}^2}{\xi_{m,n}^2} \tag{3.47}$$

and it follows in a similar fashion that

$$\frac{\sqrt{\tau^2_{1(m+1)} + \tau^2_{2n}}}{\sqrt{\tau^2_{1m} + \sigma^2_1}\sqrt{B}} = \frac{\tau^2_{1(m+1)} + \tau^2_{2n}}{\xi_{m,n}} . \tag{3.48}$$

Therefore, integral (3.38) equals

$$\frac{\tau^2_{1(m+1)} + \tau^2_{2n}}{\xi_{m,n}} \phi \left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) \tag{3.49}$$

(b) Evaluation Of The Integral

$$\int_{-\infty}^{\infty} \eta_{m+1,n}(x) \Phi\left(\frac{\eta_{m+1,n}(x)}{\xi_{m+1,n}}\right) f_G(x|\underline{x}^m)\,dx \tag{3.50}$$

This equals

$$\int_{-\infty}^{\infty} \eta_{m+1,n}(x) \left[\int_{-\infty}^{\eta_{m+1,n}(x)/\xi_{m+1,n}} \phi(t)\,dt\right] f_G(x|\underline{x}^m)\,dx \tag{3.51}$$

which is of the form

$$\int_{-\infty}^{\infty} (-ax+b) \left[\int_{-\infty}^{\frac{-ax+b}{c}} \phi(t)\,dt\right] f(x)\,dx \tag{3.52}$$

where $a = \frac{\tau^2_{1(m+1)}}{\sigma^2_1}$, $b = \rho_{2n} - \frac{\rho_{1m}\tau^2_{1(m+1)}}{\tau^2_{1m}}$, $c^2 = \tau^2_{1(m+1)} + \tau^2_{2n}$, and $f(x)$ is the Normal density $N(\rho_{1m}, \tau^2_{1m} + \sigma^2_1)$. If two consecutive changes of variable are made, firstly $u = t + ax/c$ (eliminating t), and then $y = ax/c$, and the order of integration is reversed, this equals

$$\int_{-\infty}^{b/c} \left[\int_{-\infty}^{\infty} (-cy+b)\phi(u-y)g(y)\,dy\right] du \tag{3.53}$$

where $g(y)$ is the Normal density $N(a\rho_{1m}/c,\ a^2(\tau^2_{1m} + \sigma^2_1)/c^2)$. Now it is well-known (see e.g. Feller (1966) p.45) that the convolution of two Normal density functions is a Normal density function, namely

$$\int_{-\infty}^{\infty} \phi(u-y)g(y)\,dy = h(u),\ h(u) \text{ is } N(a\rho_{1m}/c, \{a^2(\tau^2_{1m} + \sigma^2_1)/c^2\} + 1) \tag{3.54}$$

which will be abbreviated to $N(\alpha, \gamma^2)$. Furthermore,

$$h'(u) = \frac{d}{du}\int_{-\infty}^{\infty} \phi(u-y)\,g(y)\,dy$$

$$= \int_{-\infty}^{\infty} \frac{\partial}{\partial u}\,\phi(u-y)\,g(y)\,dy$$

$$= \int_{-\infty}^{\infty} (y-u)\,\phi(u-y)\,g(y)\,dy \qquad (3.55)$$

implies that

$$\int_{-\infty}^{\infty} y\phi(u-y)\,g(y)\,dy = h'(u) + uh(u)$$

$$= \frac{\alpha-u}{\gamma^2}\,h(u) + uh(u)$$

$$= \frac{(\gamma^2-1)u+\alpha}{\gamma^2}\,h(u). \qquad (3.56)$$

Therefore, integral (3.50) equals

$$-\frac{c(\gamma^2-1)}{\gamma^2}\int_{-\infty}^{b/c} uh(u)\,du + \frac{b\gamma^2-c\alpha}{\gamma^2}\int_{-\infty}^{b/c} h(u)\,du\,. \qquad (3.57)$$

Converting to standard Normal ($z = (u-\alpha)/\gamma$) gives

$$\frac{b\gamma^2-c\alpha}{\gamma^2}\int_{-\infty}^{b/c} h(u)\,du = \frac{b\gamma^2-c\alpha}{\gamma^2}\,\Phi\left(\frac{b-c\alpha}{c\gamma}\right) \qquad (3.58)$$

and

$$\frac{c(\gamma^2-1)}{\gamma^2}\int_{-\infty}^{b/c} uh(u)\,du = \frac{c(\gamma^2-1)}{\gamma^2}\left[\gamma\int_{-\infty}^{\frac{b-c\alpha}{c\gamma}} z\phi(z)\,dz + \alpha\Phi\left(\frac{b-c\alpha}{c\gamma}\right)\right]$$

$$= \frac{c(\gamma^2-1)}{\gamma^2}\left[\alpha\Phi\left(\frac{b-c\alpha}{c\gamma}\right) - \gamma\phi\left(\frac{b-c\alpha}{c\gamma}\right)\right]. \qquad (3.59)$$

Combining the last 3 equations, it follows that the integral (3.50) equals

$$(b-c\alpha)\,\Phi\left(\frac{b-c\alpha}{c\gamma}\right) + \frac{c(\gamma^2-1)}{\gamma}\,\phi\left(\frac{b-c\alpha}{c\gamma}\right) \qquad (3.60)$$

where

$$b - c\alpha = \rho_{2n} - \frac{\rho_{1m}\tau_1^2(m+1)}{\tau_{1m}^2} - \frac{\tau_1^2(m+1)\,\rho_{1m}}{\sigma_1^2}$$

$$= \rho_{2n} - \frac{\tau_{1(m+1)}^2 (\tau_{1m}^2 + \sigma_1^2)}{\tau_{1m}^2 \sigma_1^2} \rho_{1m}$$

$$= \eta_{m,n}, \text{ using (3.44)}, \tag{3.61}$$

$$(c\gamma)^2 = a^2(\tau_{1m}^2 + \sigma_1^2) + c^2$$

$$= \frac{\tau_{1(m+1)}^4 (\tau_{1m}^2 + \sigma_1^2) + \sigma_1^4 (\tau_{1(m+1)}^2 + \tau_{2n}^2)}{\sigma_1^4}$$

$$= \frac{\tau_{1(m+1)}^2 (\tau_{1m}^2 + \sigma_1^2) + \sigma_1^2 \tau_{2n}^2}{\sigma_1^2}$$

$$= \xi_{m,n}^2 , \quad \text{and} \tag{3.62}$$

$$\frac{c(\gamma^2 - 1)}{\gamma} = \frac{c^2(\gamma^2 - 1)}{c\gamma}$$

$$= \frac{a^2(\tau_{1m}^2 + \sigma_1^2)}{\xi_{m,n}} \qquad \text{(from (3.62) above)}$$

$$= \frac{\tau_{1(m+1)}^2 \tau_{1m}^2}{\sigma_1^2 \xi_{m,n}} \tag{3.63}$$

using (3.44) in each case. Therefore, combining the above 4 equations, the integral (3.50) equals

$$\eta_{m,n} \Phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) + \frac{\tau_{1(m+1)}^2 \tau_{1m}^2}{\sigma_1^2 \xi_{m,n}} \phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) . \tag{3.64}$$

Consequently, combining the results of sections (a) and (b), it follows from (3.49) and (3.64) that for any integers m and n

$$E\left[L_{m,n}^{(1)}(X_{m+1}) \,\middle|\, \underline{x}^m, \underline{y}^n\right] = \left[\frac{\tau_{1(m+1)}^2 + \tau_{2n}^2}{\xi_{m,n}} + \frac{\tau_{1(m+1)}^2 \tau_{1m}^2}{\sigma_1^2 \xi_{m,n}}\right] \phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right)$$

$$+ \eta_{m,n} \Phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right)$$

$$= \left[\frac{\tau_{1(m+1)}^2 (\tau_{1m}^2 + \sigma_1^2) + \sigma_1^2 \tau_{2n}^2}{\sigma_1^2 \xi_{m,n}}\right] \phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right)$$

$$+ \eta_{m,n} \Phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right)$$

$$= \xi_{m,n} \phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) + \eta_{m,n} \Phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) \quad \text{(using (3.44) and (3.26))}$$

$$= L^{(1)}_{m,n} \tag{3.65}$$

i.e. for each $n = 0,1,\ldots$, $\{L^{(1)}_{m,n}, \mathcal{F}_{m,n}, m \in M\}$ is a martingale.

(c) Evaluation Of The Integral

$$\int_{-\infty}^{k^{(1)}_{m+1,n}} \eta_{m+1,n}(x) f_G(x|\underline{x}^m)\, dx \tag{3.66}$$

this (cf. (3.52)) is equal to

$$\int_{-\infty}^{k^{(1)}_{m+1,n}} (-ax+b) f(x)\, dx \tag{3.67}$$

where a,b and f(x) are defined as in the previous section. This integral may be evaluated by converting to standard Normal. Letting $z = (x-\rho_{1m})/\sqrt{\tau^2_{1m}+\sigma^2_1}$, it follows that

$$-ax+b = -\frac{\tau^2_{1(m+1)}}{\sigma^2_1}\left(\sqrt{\tau^2_{1m}+\sigma^2_1}\, z + \rho_{1m}\right) + \rho_{2n} - \rho_{1m}\frac{\tau^2_{1(m+1)}}{\tau^2_{1m}}$$

$$= -\frac{\tau_{1(m+1)}\tau_{1m}}{\sigma_1} z + \eta_{m,n} \quad \text{(using (3.44))}$$

$$= -s_{1(m+1)} z + \eta_{m,n} \quad , \text{ say.} \tag{3.68}$$

Moreover,

$$x = k^{(1)}_{m+1,n} \quad \text{iff} \quad z = \frac{k^{(1)}_{m+1,n} - \rho_{1m}}{\sqrt{\tau^2_{1m}+\sigma^2_1}}$$

$$= \frac{\rho_{2n}\sigma^2_1\tau^2_{1m} - \rho_{1m}\tau^2_{1(m+1)}(\tau^2_{1m}+\sigma^2_1)}{\tau^2_{1m}\tau^2_{1(m+1)}\sqrt{\tau^2_{1m}+\sigma^2_1}} \quad \text{(from (3.34))}$$

$$= \frac{\eta_{m,n}\sigma^2_1\tau^2_{1m}}{\tau^2_{1m}\tau_{1(m+1)}\tau_{1m}\sigma_1} \quad \text{(using (3.44))}$$

$$= \frac{\eta_{m,n}}{s_{1(m+1)}} \quad . \tag{3.69}$$

Therefore, the integral (3.66) equals

$$- s_{1(m+1)} \int_{-\infty}^{\eta_{m,n}/s_{1(m+1)}} z\phi(z)\,dz \quad + \quad \eta_{m,n} \int_{-\infty}^{\eta_{m,n}/s_{1(m+1)}} \phi(z)\,dz$$

$$= s_{1(m+1)}\ \phi\left(\frac{\eta_{m,n}}{s_{1(m+1)}}\right) + \eta_{m,n}\ \Phi\left(\frac{\eta_{m,n}}{s_{1(m+1)}}\right) . \tag{3.70}$$

Thus the expected loss plus cost of observing X_{m+1} and then stopping is

$$E\left\{U_{m+1,n}\left[g_{m,n}(X_{m+1})\right] \middle| \underline{x}^m, \underline{y}^n\right\} = \eta_{m,n}\left\{\Phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) - \Phi\left(\frac{\eta_{m,n}}{s_{1(m+1)}}\right)\right\}$$

$$+ \xi_{m,n}\ \phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) - s_{1(m+1)}\ \phi\left(\frac{\eta_{m,n}}{s_{1(m+1)}}\right) + (m+n+1)c . \tag{3.71}$$

By a symmetrical argument the expected loss plus cost of observing Y_{n+1} and then stopping is

$$E\left\{U_{m,n+1}\left[g_{m,n}(Y_{n+1})\right] \middle| \underline{x}^m, \underline{y}^n\right\} = \eta_{m,n}\left\{\Phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) - \Phi\left(\frac{\eta_{m,n}}{s_{2(n+1)}}\right)\right\}$$

$$+ \xi_{m,n}\ \phi\left(\frac{\eta_{m,n}}{\xi_{m,n}}\right) - s_{2(n+1)}\ \phi\left(\frac{\eta_{m,n}}{s_{2(n+1)}}\right) + (m+n+1)c . \tag{3.72}$$

To find the minimum of these two expressions, consider the function

$$h(\eta,s) = \eta\Phi\left(\frac{\eta}{s}\right) + s\phi\left(\frac{\eta}{s}\right) , \quad -\infty < \eta < \infty, \quad 0 < s < \infty . \tag{3.73}$$

Firstly, note that

$$\frac{\partial h}{\partial s} = -\frac{\eta^2}{s^2}\phi\left(\frac{\eta}{s}\right) + \phi\left(\frac{\eta}{s}\right) + \frac{\eta^2}{s^2}\phi\left(\frac{\eta}{s}\right)$$

$$= \phi\left(\frac{\eta}{s}\right)$$

$$> 0 \quad \text{for all } s > 0, \text{ for any } \eta \tag{3.74}$$

i.e. h is increasing in s for any value of η. Recalling the definitions

$$s^2_{1(m+1)} = \frac{\tau^2_{1(m+1)}\tau^2_{1m}}{\sigma_1^2} , \quad s^2_{2(n+1)} = \frac{\tau^2_{2(n+1)}\tau^2_{2n}}{\sigma_2^2} \tag{3.75}$$

it follows that if $\sigma_1^2 = \sigma_2^2 = \sigma^2$, say, then

$$\tau_{1m}^2 > \tau_{2n}^2 \quad \text{implies} \quad \frac{\sigma^2\beta_1^2}{m\beta_1^2+\sigma^2} > \frac{\sigma^2\beta_2^2}{n\beta_2^2+\sigma^2}$$

$$\text{implies} \quad n\sigma^2\beta_1^2\beta_2^2 + \sigma^4\beta_1^2 > m\sigma^2\beta_1^2\beta_2^2 + \sigma^4\beta_2^2$$

$$\text{implies} \quad (n+1)\sigma^2\beta_1^2\beta_2^2 + \sigma^4\beta_1^2 > (m+1)\sigma^2\beta_1^2\beta_2^2 + \sigma^4\beta_2^2$$

$$\text{implies} \quad \tau_{1(m+1)}^2 > \tau_{2(n+1)}^2$$

i.e. $\tau_{1m}^2 > \tau_{2n}^2$ implies $\tau_{1m}^2 > \tau_{2n}^2$ and $\tau_{1(m+1)}^2 > \tau_{2(n+1)}^2$

$$\text{implies} \quad s_{1(m+1)}^2 > s_{2(n+1)}^2 \ . \tag{3.76}$$

Thus if $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then each s^2 (and hence each s) increases with its associated τ^2. Consequently, the minimum expected loss plus cost of taking one more observation in an optimal fashion and then stopping, given that experimentation is at the (m,n)-th stage, is

$$V_{m+n+1}^{M\,(*)}(\underline{x}^m,\underline{y}^n) = \begin{cases} E\left\{U_{m+1,n}\left[g_{m,n}(X_{m+1})\right] \middle| \underline{x}^m,\underline{y}^n\right\} & \text{if } s_{1(m+1)} > s_{2(n+1)} \\ \text{either} & \text{if } s_{1(m+1)} = s_{2(n+1)} \\ E\left\{U_{m,n+1}\left[g_{m,n}(Y_{n+1})\right] \middle| \underline{x}^m,\underline{y}^n\right\} & \text{if } s_{1(m+1)} < s_{2(n+1)} \end{cases} \tag{3.77}$$

where, if $\sigma_1^2 = \sigma_2^2$, the s's may be replaced by the corresponding posterior variances. (cf. (2.57)). Therefore, the solution of the last component $\phi_{M,N}^{(M+N+1)*}$ of the optimal stopping rule for the problem truncated after $M+N+1$ observations is possible. Note that

$$\begin{aligned} U_{M,N}(g_{M,N}) - V_{M+N+1}^{M\,(*)}(\underline{x}^M,\underline{y}^N) &= Q(\eta_{M,N},\xi_{M,N}) - \min_{\substack{K=M,N \\ (i=1,2)}} \left\{h(\eta_{M,N},\xi_{M,N}) - h(\eta_{M,N},s_{i(K+1)})\right\} \\ &\qquad - c \\ &= \max\left\{Q(\eta_{M,N},s_{1(M+1)}),\ Q(\eta_{M,N},s_{2(N+1)})\right\} - c \end{aligned} \tag{3.78}$$

Therefore

$$\phi_{M,N}^{(M+N+1)*}(\underline{x}^M,\underline{y}^N) = \begin{cases} 1 & \text{if } \max\left\{Q(\eta_{M,N,s_{1(M+1)}}), Q(\eta_{M,N,s_{2(N+1)}})\right\} < c \\ \text{any} & \text{if } = \\ 0 & \text{if } > \end{cases} \tag{3.79}$$

(4) Stopping Boundaries In (m,n)-Space; Motivation For An Alternative Sequential Design Procedure

From the preceding two sections it is clear that the computation of the Bayes sequential design rule for the truncated problem is difficult and laborious; in this section consideration of the stopping boundaries in terms of (m,n) leads to a sequential design procedure for both truncated and non-truncated problems, which, although obviously not optimal, is very easy to apply and is clearly superior to the usual decision rule (pairwise (x_i,y_i) design).

Consider the function $Q(\eta,s)$ defined by (3.30). Clearly, for any $s > 0$, $Q(\eta,s)$ is symmetric about $\eta = 0$, and hence

$$\left.\frac{\partial Q}{\partial \eta}\right|_{\eta=a} = -\left.\frac{\partial Q}{\partial \eta}\right|_{\eta=-a} \quad \text{for any } a > 0 . \tag{3.80}$$

In fact, for $\eta < 0$,

$$\frac{\partial Q}{\partial \eta} = \Phi\left(\frac{\eta}{s}\right) + \frac{\eta}{s}\phi\left(\frac{\eta}{s}\right) - \frac{\eta}{s}\phi\left(\frac{\eta}{s}\right)$$

$$= \Phi\left(\frac{\eta}{s}\right) \tag{3.81}$$

which is positive-valued for all $\eta < 0$, for any value of $s > 0$. Thus $\frac{\partial Q}{\partial \eta}$ for $\eta > 0$ equals $-\Phi\left(\frac{-\eta}{s}\right)$ which is negative-valued for all $\eta > 0$, for any value of $s > 0$. Consequently, for a given $s > 0$, $Q(\eta,s) \equiv Q_s(\eta)$ is shown in Figure 3.1 below.

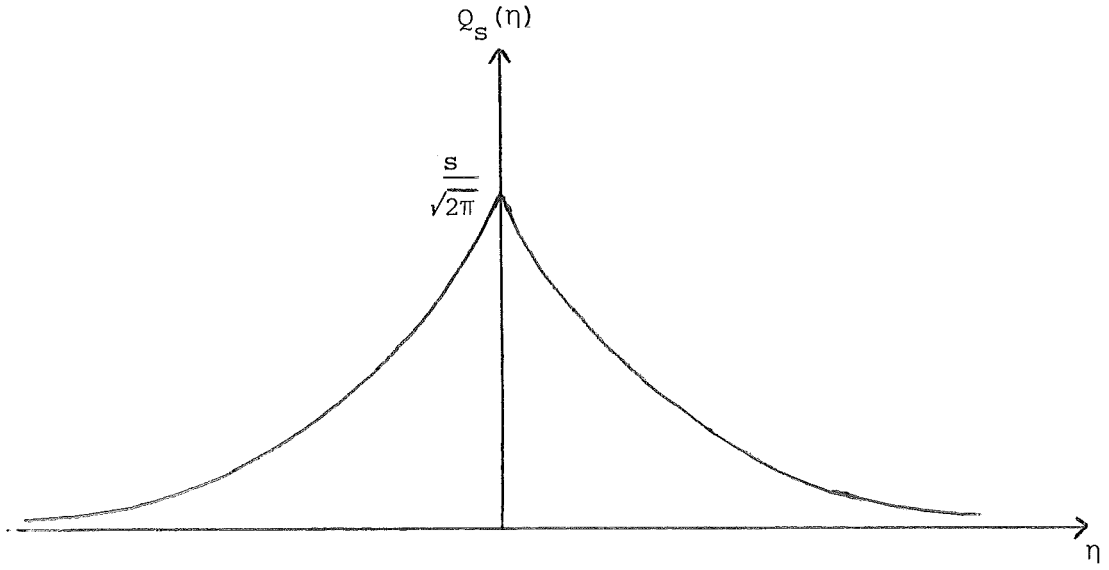

FIGURE 3.1. The function $Q_s(\eta)$ for a given value of s.

Given s, $Q_s(\eta)$ has a maximum value at $\eta = 0$, i.e.

$$Q_s(\eta) \leqslant \frac{s}{\sqrt{2\pi}} \qquad \text{for all } \eta \tag{3.82}$$

and

$$\frac{\partial Q}{\partial s} = \frac{\partial h}{\partial s} \qquad \text{(comparing (3.30) with (3.73))}$$

$$= \phi\left(\frac{\eta}{s}\right) \tag{3.83}$$

which is positive-valued for all $s > 0$, for any η; i.e. Q increases in s for any η. Furthermore, it follows from (3.75) that each s increases in its corresponding τ, which decreases in its corresponding sample size. in other words,

$$s_{1m} > s_{1(m+1)} > s_{1(m+2)} \quad \ldots \quad \text{for any positive integer m}$$

$$s_{2n} > s_{2(n+1)} > s_{2(n+2)} \quad \ldots \quad \text{for any positive integer n .} \tag{3.84}$$

Therefore, (3.83) and (3.84) together imply that $Q(\eta, s_{1(m+1)})$ decreases with m for any fixed value of η, and $Q(\eta, s_{2(n+1)})$ decreases with n for any fixed value of η. In particular, of course, this is true for $\eta = 0$. Therefore, defining the integers m_0 and n_0 by

$$m_0 = \text{largest integer m such that } \frac{s_{1(m+1)}}{\sqrt{2\pi}} \geqslant c$$

$$n_0 = \text{largest integer n such that } \frac{s_{2(n+1)}}{\sqrt{2\pi}} \geqslant c \qquad (3.85)$$

it follows that it is always optimal to stop at stage (m,n) (if stopping has not already occurred) if $m > m_0$ and $n > n_0$; see Figure 3.2.

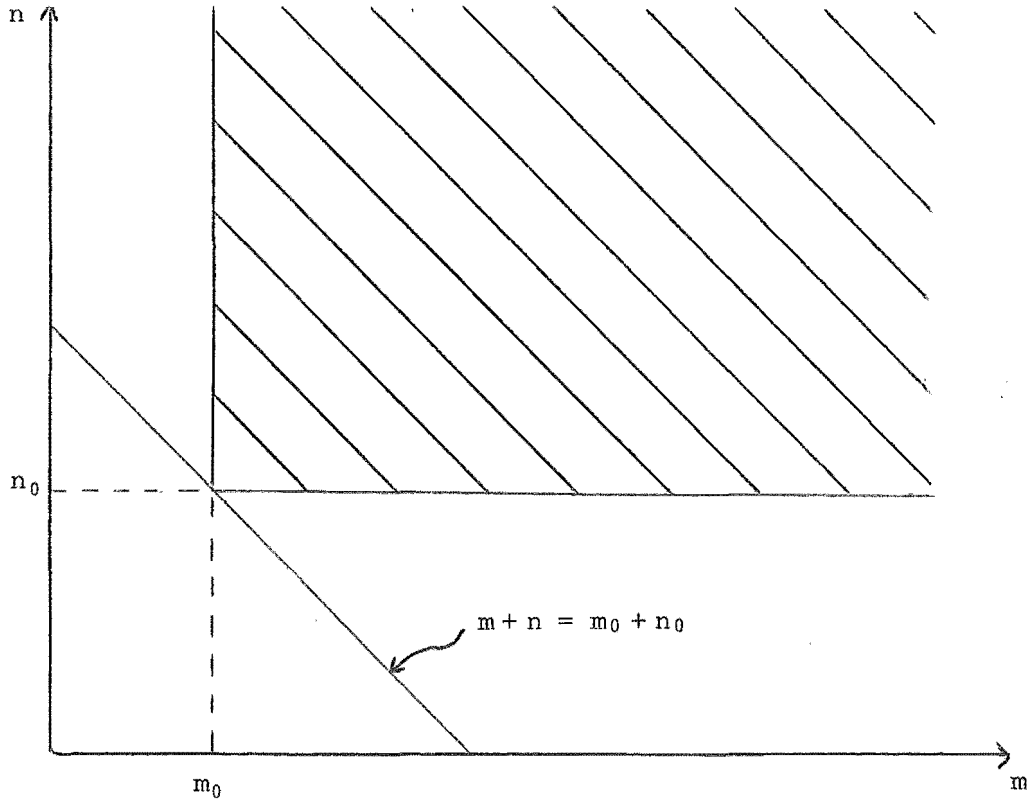

FIGURE 3.2. Stopping region in (m,n)-space.

In other words, if sampling proceeds into the shaded region, it is then optimal to stop experimentation, regardless of whether the problem is truncated or non-truncated. This is true for two reasons: firstly, if the problem is truncated after M+N+1 observations and experimentation is at (M,N) inside the stopping region, then the inequalities in (3.85) are reversed and consequently (see (3.79)) the optimal stopping rule will stop; and secondly, if the problem is truncated at a larger value than M+N+1, or non-truncated, it is optimal to stop from Theorem 3.1 below.

Theorem 3.1. In the notation of Chapter 2, if $r(G, (\underline{\phi}^{J*}, \underline{\gamma}^{(J-1)*}, \underline{\delta}^{J*})) \to r(G, (\underline{\phi}^{*}, \underline{\gamma}^{*}, \underline{\delta}^{*}))$ as $J \to \infty$, and if for all $j > J_0$

$$\rho_{j-1}(g_{j-1}) - E_{\gamma_{j-1}}\{\rho_j[g_{j-1}(X_j,E_j)] \mid \mathcal{F}_{j-1}\}$$

$$\leqslant E_{\gamma_{j-1}}\{c_j[g_{j-1}(X_j,E_j)] \mid \mathcal{F}_{j-1}\} - c_{j-1}(g_{j-1}) \quad (\text{a.s. } \mathcal{F}_{j-1})$$

$$\text{for all experiment rules } \gamma_{j-1} \tag{3.86}$$

then $r(G, (\underline{\phi}^{J_0*}, \underline{\gamma}^{(J_0-1)*}, \underline{\delta}^{J_0*})) = r(G, (\underline{\phi}^{*}, \underline{\gamma}^{*}, \underline{\delta}^{*}))$.

Proof: This follows as an obvious generalisation of Theorem 4 in Section 7.2 of Ferguson (1967). □

The situation described by (3.86) is, perhaps, the "sequential design" analogue of the Monotone Case in sequential decision theory: see Chow, Robbins and Siegmund (1971), Section 3.5. However, in this example, (3.86) does not hold for every (m,n) satisfying $m+n = m_0+n_0$; only for $(m,n) = (m_0, n_0)$ (see Figure 3.2). Moreover, because the two "continuation" regions in Figure 3.2 are unbounded in either one of m or n, it follows that there is no J_0 for which (3.86) holds for all (m,n) satisfying $m+n = J_0$. Therefore, the monotonicity property (3.86) in the design case depends on both the sample size and the experimentation history.

Nevertheless, if the risk at (m_0, n_0) is acceptable to the statistician (note that the risk is bounded by $\xi_{m_0,n_0}/\sqrt{2\pi} + (m_0+n_0)c$, from (3.31) and (3.82)), then the fixed sample size design rule, which specifies m_0 observations from Π_1 and n_0 observations from Π_2 followed by the Bayes terminal decision, involves no complicated backward induction computations and will in many cases yield a risk and expected sample size which are not much larger than the corresponding values for the optimal sequential rule. Of course, the stopping region in Figure 3.2 is a "smallest possible" region; there may be sets $\{(\underline{x}^n, \underline{e}^n) : n < m_0 + n_0\}$ for which the optimal sequential rule will stop. As an example of this, note that

$$\tau^2_{1(m+1)} < \sigma_1^2 \quad \text{and} \quad \tau^2_{2(n+1)} < \sigma_2^2$$

(since β_1^2 and β_2^2 are both finite)

$$\text{implies} \quad \tau^2_{1(m+1)} - \sigma_1^2 < 0 \quad \text{and} \quad \tau^2_{2(n+1)} - \sigma_2^2 < 0$$

$$\text{implies} \quad \frac{\tau^2_{1m}(\tau^2_{1(m+1)} - \sigma_2^2)}{\sigma_1^2} < 0 < \tau^2_{2n}$$

$$\text{and} \quad \frac{\tau^2_{2n}(\tau^2_{2(n+1)} - \sigma_2^2)}{\sigma_2^2} < 0 < \tau^2_{1m}$$

$$\text{implies} \quad s_{1(m+1)} < \xi_{m,n} \quad \text{and} \quad s_{2(n+1)} < \xi_{m,n} \,. \tag{3.87}$$

It follows from this and (3.83) that, for any value $\eta_{m,n}$,

$$Q(\eta_{m,n}, \xi_{m,n}) > \max\{Q(\eta_{m,n}, s_{1(m+1)}), Q(\eta_{m,n}, s_{2(n+1)})\} \,. \tag{3.88}$$

Therefore, if $(m,n) = (m_0 - 1, n_0)$ or $(m_0, n_0 - 1)$ and $Q(\eta_{m,n}, \xi_{m,n}) \leqslant c$, it is clearly optimal to stop by (3.79), at a cost of $m_0 + n_0 - 1$ observations. However, because the optimal sequential rule is solvable for the first stage of backward induction only, it is not possible to characterise the optimal stopping region in $\mathcal{X}^m \times \mathcal{Y}^n$ for $\{(m,n) : m < m_0 \text{ and } n < n_0\}$.

Despite this, it seems clear from the considerations above that if m_0 and n_0 are some distance apart, then the fixed sample size design rule will be superior to any fixed sample size decision rule which specifies that at least $\max\{m_0, n_0\}$ observations are to be taken from each population. Table 3.3 lists some values of σ, β and k, the sample size, for which

$$\frac{s_{k+1}}{\sqrt{2\pi}} = c \tag{3.89}$$

for $c = \frac{1}{10}$ and $\frac{1}{20}$, where $s_{k+1} = \tau_k \tau_{k+1}/\sigma$ and $\tau_k^2 = \sigma^2\beta^2/(k\beta^2 + \sigma^2)$.

TABLE 3.3. Values of σ, β, c and k for which (3.89) is true.

σ	β	k ($c = .1$)	k ($c = .05$)
1	1	3	7
1	2	4	8
2	1	4	12
1	3	4	8
2	3	8	15
1	4	4	8
2	4	8	16
3	4	11	23
2	5	8	16
3	5	12	24

4. CONCLUSIONS

In this chapter an attempt has been made to apply the constructive theory of Bayes sequential design procedures to some typical sequential design problems, namely an identification problem and a ranking problem.

Whilst the solution of the simple identification problem illustrates the power of incorporating the design aspect into the sequential rule, it becomes clear that the recursive computation of the $V_j^{(\gamma)}$-functions in a seemingly well-behaved ranking problem such as the two-normal/normal priors example is not often possible (at least not analytically ; the necessity of numerical integration techniques defeats the purpose of employing a sequential design procedure in the first place). In fact, this is often the case with Bayes sequential design procedures.

From a practical point of view, then, it is necessary either to investigate situations (such as the monotone case) in which the optimal design procedure is of a somewhat simpler form (i.e. not involving recursive backward induction), or to consider the performance of procedures which, although not optimal, are nevertheless easy to apply in practice and sufficiently good (in terms of risk and expected sample size) to be worth using. At the very least, it is surely of interest to investigate procedures whose performance appears to be better than those currently in use.

CHAPTER IV

COMPUTATIONAL CRITERIA: THE ONE-STEP-AHEAD EXPERIMENT RULE FOR A MODIFIED GOAL

1. THE ONE-STEP-AHEAD EXPERIMENT RULE AND ITS PERFORMANCE

As was seen in Chapter 3, the optimal pair $(\underline{\phi}^*, \underline{\gamma}^*)$ comprising the optimal stopping rule and the optimal experiment rule, although defined constructively in theory for the truncated case, is extremely difficult if not impossible to compute in most situations of interest. This is even more so in the design case than in the non-design case, because not only are the optimal continuation risks defined in (2.61) difficult to compute, but also the number of possible allocations of the n observations over the k available experiments (i.e. populations) increases exponentially with n. Furthermore, this is without regard to the possible outcomes of the random variables themselves. In the simplest case of k binomial populations considered in the previous chapter, where each random variable is restricted to two possible outcomes 0 and 1, the number of possible values of the sufficient statistic t_n based on the outcomes of n random variables is $\binom{n + 2k - 1}{2k - 1}$. For example, if $n = 4$ and $k = 2$, this gives 35 possible values of t_n, as opposed to 9 for the corresponding decision problem.

Therefore, it seems desirable to investigate the behaviour of rules $(\underline{\phi}, \underline{\gamma})$ which in some sense "perform well" when compared to the optimal rule $(\underline{\phi}^*, \underline{\gamma}^*)$ yet are relatively easy to compute. The criteria for performance used is the Bayes risk of the procedure and its expected sample size. If it is possible to find a rule $(\underline{\phi}^0, \underline{\gamma}^0)$ whose Bayes risk and expected sample size are not much greater than those of $(\underline{\phi}^*, \underline{\gamma}^*)$, but which can be used for moderately large sample sizes with little computational effort, then it is possible to solve a large range of

problems easily and efficiently. In fact, if the risk of such a procedure is "close to" the risk of the Bayes decision procedure, which observes a r.v. from each population at each stage, but the expected sample size is significantly smaller, this constitutes an improvement in multiple decision problems. This chapter is devoted to the study of the following procedure.

The One-Step-Ahead or Myopic Experiment Rule

At each stage of experimentation, select the next experiment as if there were only one stage of experimentation left. With regard to experiment selection only, then, the risk from taking one more observation and then proceeding optimally is equivalent to the risk of taking one more observation, i.e.

$$E_{\gamma_{j-1}}\left\{V_j^{(\gamma)}[g_{j-1}(X_j,E_j)]\,|\,F_{j-1}\right\} \equiv E_{\gamma_{j-1}}\left\{U_j[g_{j-1}(X_j,E_j)]\,|\,F_{j-1}\right\} \tag{4.1}$$

where γ_{j-1}^0 is defined according to

$$M_{V_j^{(\gamma^0)}}(\underline{x}^{j-1},\underline{e}^{j-1}) = \operatorname*{ess\,inf}_{e_j\in\mathcal{E}} E\left\{U_j[g_{j-1}(X_j,e_j)]\,|\,G_{j-1}\right\} \tag{4.2}$$

$(j = 1,2,\ldots)$.

The One-Step-Ahead or Myopic Stopping Rule

At each stage of experimentation, stop if the current risk is less than the expected risk of taking one more observation and then stopping (i.e. decide whether to stop as if there were only one stage of experimentation left). Then, with regard to stopping only (for a given experiment rule γ) (4.1) may be regarded as true, and each ϕ_{j-1}^0 is defined by

$$\phi_{j-1}^0(\underline{x}^{j-1},\underline{e}^{j-1}) = \begin{cases} 1 & \text{if } U_{j-1}(g_{j-1}) < E_{\gamma_{j-1}}\left\{U_j[g_{j-1}(X_j,E_j)]\,|\,F_{j-1}\right\} \\ \text{any} & \text{if } = \\ 0 & \text{if } > \end{cases} \tag{4.3}$$

$(j = 1,2,\ldots)$.

(4.1) is expressed rather loosely (it not being clear how the V_j's are defined) and is intended rather to highlight the obvious; that a sequential design procedure incorporating either of the one-step-ahead rules $\underline{\phi}^0$ and $\underline{\gamma}^0$ will not, in general, be optimal. A one-step-ahead rule (henceforth abbreviated to OSA) ignores all potential information available from possible future stages of experimentation, and in particular a OSA stopping rule will always stop no later and probably much earlier than the optimal stopping rule, from (2.51). However, a OSA rule is clearly much easier to compute than the optimal rule, and does not depend on whether the problem is truncated or not. Therefore, it is of interest to consider the performance of $(\underline{\phi}^0, \underline{\gamma}^0)$ or $(\underline{\phi}^*, \underline{\gamma}^0)$ relative to $(\underline{\phi}^*, \underline{\gamma}^*)$, and in particular, under what conditions a OSA rule is optimal.

(1) Clearly $(\underline{\phi}^0, \underline{\gamma}^0)$ will be optimal after J_0 stages of experimentation if, for all $j > J_0$,

$$E_{\gamma^*_{j-1}}\left\{V_j^{(\underline{\gamma}^*)}[g_{j-1}(X_j,E_j)]\,|\,F_{j-1}\right\} = E_{\gamma^0_{j-1}}\left\{U_j[g_{j-1}(X_j,E_j)]\,|\,F_{j-1}\right\} \qquad (4.4)$$

i.e. (4.1) is true for all $j > J_0$.* However, it follows from the definition of each $V_j^{(.)}$ that $V_j^{(.)} \leqslant U_j$ for each $j = 0,1,\ldots$, and hence (4.4) implies that $V_j^{(\underline{\gamma}^*)} = U_j$ for all $j > J_0$. Thus the optimal stopping rule is to stop after $J_0 + 1$ observations if stopping has not already occurred, and hence the knowledge that all but the first J_0 components of $\underline{\phi}^*$ and $\underline{\gamma}^*$ are OSA is of little use; it is the first J_0 components of each rule which are needed. Consequently, if (4.4) is true for all $j = 0,1,\ldots$, i.e. $(\underline{\phi}^0, \underline{\gamma}^0)$ is equivalent to $(\underline{\phi}^*, \underline{\gamma}^*)$, then the optimal rule will take at most one observation and then stop. It follows that $(\underline{\phi}^0, \underline{\gamma}^0)$ is optimal in the sense of (4.4) only in the trivial case when at most one observation is taken.

*Unless specified otherwise, equations such as (4.4) refer to both truncated and non-truncated problems.

(2) From the definition of the optimal experiment rule $\underline{\gamma}^*$, it follows that $(\underline{\phi}^*, \underline{\gamma}^0)$ is optimal if the following condition is true for each $j = 1,2,\ldots$ (possibly truncated):

$$e^* \text{ attains } \operatorname*{ess\,inf}_{e \in \mathcal{E}} E\left\{V_j^{(\underline{\gamma}^*)}[g_{j-1}(X_j,e)] \,|\, G_{j-1}\right\}$$

$$\text{iff } e^* \text{ attains } \operatorname*{ess\,inf}_{e \in \mathcal{E}} E\left\{U_j[g_{j-1}(X_j,e)] \,|\, G_{j-1}\right\} \qquad (4.5)$$

i.e. the optimal "path" of experimentation coincides with the "path" of experimentation chosen according to $\underline{\gamma}^0$. However, there appears to be no way of verifying the condition (4.5) without resorting to computation of the $V_j^{(\cdot)}$ functions, which negates the purpose of considering the OSA rule; further, if it were possible to characterise situations where (4.5) was true, it would still be necessary to compute the $V_j^{(\cdot)}$ functions in order to solve for $\underline{\phi}^*$. Therefore, it may be concluded that the characterisation of optimality conditions for even a partial OSA rule is virtually impossible except in the trivial case (1).

The performance of the OSA rule $(\underline{\phi}^0, \underline{\gamma}^0)$ has been compared to that of the Bayes sequential design procedure and the Bayes sequential decision procedure for the binomial identification problem considered in Chapter 3. As expected, it was found that the OSA procedure results in a smaller expected sample size and a larger risk than the Bayes sequential design procedure; moreover, these differences increase as J, the truncation value, increases. Some values of these performance characteristics are listed in Table 4.1 below.

TABLE 4.1. Values of the risk and expected sample size for a) OSA, b) Bayes design, and c) Bayes decision rules (2 populations) ($\underline{g} = (\cdot 7, \cdot 3)$)

J	R	ν_1	ν_2	OSA		Design		Non-Design	
				Risk	E[N]	$V_0(\underline{\gamma}^{(J-1)*})$	E[N]	$V_0^{(J-1)}$	E[N]
4	50	·2	·7	9.3570	1.539	8.4489	2.933	8.8140	3.180
6	50	·2	·7	8.8488	1.566	7.6019	3.467	7.8571	3.855
8	50	·2	·7	8.8372	1.569	7.2118	3.753	7.3388	4.238

From these and other results considered, it appears that the risk and expected sample size of the OSA procedure are both relatively unchanged for sufficiently large values of J, i.e. little is gained or lost in terms of performance for larger values of the truncation point J. This is perhaps to be expected, as the OSA rule is a "forward-looking" rule rather than a backward induction rule. Mathematically speaking, a rule $(\underline{\phi}, \underline{\gamma})$ may be thought of as "forward looking" if for any integers J_1 and J_2 with $J_1 < J_2$,

$$\underline{\phi}^{J_2} = (\underline{\phi}^{J_1}, \phi_{J_1+1}, \ldots, \phi_{J_2})$$
$$\underline{\gamma}^{J_2} = (\underline{\gamma}^{J_1}, \gamma_{J_1+1}, \ldots, \gamma_{J_2}) \qquad (4.6)$$

i.e. the rule for the problem truncated at J_1 is the "initial segment" of the rule for the problem truncated at J_2 for any $J_2 > J_1$. Thus a "forward-looking" rule may be thought of as being independent of the truncation value; clearly the backward induction type of rule does not satisfy this condition. It follows that if $(\underline{\phi}, \underline{\gamma})$ satisfies (4.6), then in the above notation

$$P_{G,f_\theta,\underline{\phi}^{J_1},\underline{\gamma}^{J_1}}\{N = J_1\} = P_{G,f_\theta,\underline{\phi}^{J_2},\underline{\gamma}^{J_2}}\{N \geqslant J_1\} \tag{4.7}$$

for any $J_2 > J_1$. In other words, for a sufficiently small value of this probability (guaranteed by the finiteness property (2.67) of N), it is possible to find an integer J_1 such that the risk and expected sample size of a rule satisfying (4.6) will be relatively unchanged for values of J_2 greater than J_1.

Therefore, it may be concluded that the performance of the OSA rule will not be particularly good for large truncation values, which might be regarded as the type of problem for which an easily computed rule would be of most benefit. As mentioned before, if the statistician's goal were such that a certain level of risk was acceptable, and if it were known that the risks of the three procedures considered were at that level, then it might be hoped that the expected sample size of the OSA rule would be close to that of the Bayes sequential design rule, and less than that of the Bayes sequential decision rule.

Furthermore, the difficulty of computing and identifying the Bayes sequential design procedure $(\underline{\phi}^*, \underline{\gamma}^*)$ lies in the fact that $\underline{\phi}^*$ and $\underline{\gamma}^*$ are interdependent, as both depend on the solution of backward induction equations. If it were possible to modify the problem or the goal of the statistician so that $\underline{\phi}^*$ did not depend on $\underline{\gamma}^*$ or on backward induction techniques, and was easily computed, then it would be possible to consider $\underline{\gamma}^*$ alone, and how well $\underline{\gamma}^0$ would perform as an acceptable experiment rule. In other words, the problem would be reduced to a comparison of experiment rules only.

In the next section, the goal of the statistician is modified appropriately so that the stopping rule is "independent" of the experiment rule, and hence like the terminal decision rule, is of the same form for any experiment rule.

2. A MODIFIED GOAL

It is henceforth assumed (unless stated otherwise) that $A = \{a_1, a_2, \ldots, a_k\}$ and $\mathcal{A} = \{1,2,\ldots, k\}$ for some specified integer k, and that the problem is unbounded; i.e. there is no truncation value J. Finally, it is assumed that the loss function is 0 - 1 loss as defined in (3.3).

The first modification is that the cost function $c_j(\theta, \underline{x}^j, \underline{e}^j)$ is identically equal to zero for all values of the arguments, for any value of $j = 0,1,\ldots$. Theorem 4.1 then follows.

Theorem 4.1 For each a_i, $i = 1,2,\ldots, k$, define

$$U_n(a_i, g_n) = \int_{\Theta} L(\theta, a_i) g(\theta | \underline{x}^n, \underline{e}^n) d\nu(\theta) \tag{4.8}$$

to be the expected posterior loss if a_i is chosen, given $(\underline{X}^n, \underline{E}^n) = (\underline{x}^n, \underline{e}^n)$, and

$$P(\Theta_i | \underline{x}^n, \underline{e}^n) = \int_{\Theta_i} g(\theta | \underline{x}^n, \underline{e}^n) d\nu(\theta) \tag{4.9}$$

to be the posterior probability that $\theta \in \Theta_i$, given $(\underline{X}^n, \underline{E}^n) = (\underline{x}^n, \underline{e}^n)$, for any $n = 0,1,\ldots$. Then under the assumption that the loss function satisfies (3.3),

$$U_n(a_i, g_n) = 1 - P(\Theta_i | \underline{x}^n, \underline{e}^n) . \tag{4.10}$$

Proof: From (3.3)

$$\begin{aligned} \int_{\Theta} L(\theta, a_i) g(\theta | \underline{x}^n, \underline{e}^n) d\nu(\theta) &= \int_{\Theta/\Theta_i} g(\theta | \underline{x}^n, \underline{e}^n) d\nu(\theta) \\ &= \int_{\Theta} g(\theta | \underline{x}^n, \underline{e}^n) d\nu(\theta) - \int_{\Theta_i} g(\theta | \underline{x}^n, \underline{e}^n) d\nu(\theta) \\ &= 1 - \int_{\Theta_i} g(\theta | \underline{x}^n, \underline{e}^n) d\nu(\theta) \end{aligned}$$

□

Corollary 4.1 $U_n(g_n) = 1 - \max_{i=1,\ldots,k} P(\theta_i | \underline{x}^n, \underline{e}^n)$.

The interpretation of Theorem 4.1 and its corollary is that the Bayes terminal decision rule based on n observations $(\underline{x}^n, \underline{e}^n)$ selects the subspace θ_i of Θ with the largest posterior probability, given n observations $(\underline{x}^n, \underline{e}^n)$.

In order to prove the next result, the notation of Chapter 2 is used with some slight modification. Let $\mathcal{F} = \mathcal{B}(\theta, \underline{X}, \underline{E}, A)$ be the σ-field generated by the elements of a sequential design problem with no stopping rule; then the probability measure $P_{G,f_\theta,\underline{\gamma},\underline{\delta}}(\theta, \underline{x}, \underline{e}, a)$ may be defined in an analogous fashion to that used in Chapter 2. Similarly, let $\mathcal{F}_n = \mathcal{B}(\underline{X}^n, \underline{E}^n)$ for $n = 0,1,\ldots$. Then $\mathcal{F}_n \subset \mathcal{F}_m \subset \mathcal{F}$ for all integers m,n such that $m > n$, and it is well-known that for any real r.v. z,

$$\int_\Theta z\, g(\theta | \underline{x}^n, \underline{e}^n)\, d\nu(\theta)$$

is a version of the conditional expectation $E[z|\mathcal{F}_n]$. If $z = z(a)$ is chosen to be the loss function $L(\theta,a)$, then for any $a \in A$

$$U_n(a,g_n) = E[z(a)|\mathcal{F}_n]$$

and

$$U_n(g_n) = \operatorname*{ess\,inf}_{a \in A} U_n(a,g_n) \tag{4.11}$$

for $n = 0,1,\ldots$, where (4.8) may be extended to the case where A may not be finite. If the dependence on g_n is suppressed for notational convenience, then the following theorem may be stated.

Theorem 4.2 If the cost c_j is zero-valued for all j, then for any integers m,n satisfying $m > n$,

$$U_n \geqslant E[U_m|\mathcal{F}_n] \tag{4.12}$$

i.e. $\{U_n, \mathcal{F}_n, n = 0,1,\ldots\}$ is a supermartingale.

Proof: It is first shown that, for any $m > n$

$$E[U_m|\mathcal{F}_n] \leq \operatorname*{ess\,inf}_{a \in A} E[U_m(a)|\mathcal{F}_n] \quad . \tag{4.13}$$

As noted in Chapter 2, U_m is $\mathcal{F}$-measurable and integrable for each integer m. It follows from the definition of U_m that

$$U_m \leq U_m(a) \quad \text{for all } a \in A \qquad (\text{a.s. } \mathcal{F}_m)$$

implies

$$E[U_m|\mathcal{F}_n] \leq E[U_m(a)|\mathcal{F}_n] \quad \text{for all } a \in A, \ (n < m)$$

implies

$$E[U_m|\mathcal{F}_n] \leq \operatorname*{ess\,inf}_{a \in A} E[U_m(a)|\mathcal{F}_n] \qquad (n < m) \ . \tag{4.14}$$

Furthermore, it follows from the definition of $U_n(a)$ in (4.11) that $\{U_n(a), \mathcal{F}_n, n = 0,1,\ldots\}$ is a martingale for any $a \in A$ (see Chow, Robbins, and Siegmund (1971) p.12, example (a)). Therefore, for any integers m,n with $m > n$,

$$\begin{aligned} E[U_m|\mathcal{F}_n] &\leq \operatorname*{ess\,inf}_{a \in A} E[U_m(a)|\mathcal{F}_n] && \text{(proved above)} \\ &= \operatorname*{ess\,inf}_{a \in A} U_n(a) && \text{(by the martingale property)} \\ &= U_n \ . \end{aligned}$$

□

Consider the sequential design problem truncated at J observations. Under the assumption that the cost is always zero, it follows from Theorem 4.2 that the optimal stopping rule will not stop until J observations have been taken with positive probability, for any experiment rule $\underline{\gamma}^{J-1}$. Consequently, no optimal stopping rule exists for the unbounded problem. However, by modifying the statistician's goal, it may be shown that an

essentially complete class of rules $(\underline{\phi}^0, \underline{\gamma}^0)$ in the sense of (2.87) exists for the non-truncated problem, where $\underline{\phi}^0$ does not depend on backward induction for its solution. Define

$$\Gamma' = \{\underline{\gamma}: \ U_n(g_n) \to 0 \quad (a.s.) \quad \text{as } n \to \infty\} \tag{4.15}$$

to be the class of all experiment rules $\underline{\gamma}$ for the non-truncated problem for which the Bayes posterior expected loss approaches zero as n, the sample size, approaches infinity. Note that $U_n(g_n)$ is independent of $\underline{\gamma}$ only when $(\underline{X}^n, \underline{E}^n) = (\underline{x}^n, \underline{e}^n)$ is known; to determine whether the event described in (4.15) will occur or not before sampling takes place will depend on which $\underline{\gamma}$ is used.

Because $\underline{\gamma}' \in \Gamma'$ ensures that $U_n(g_n)$ converges to zero for all values $(\underline{x}^n, \underline{e}^n)$, it follows that the stopping rule $\underline{\phi}^\varepsilon$ defined for some $\underline{\gamma}' \in \Gamma'$ by

$$\begin{aligned} \phi_n^\varepsilon &\equiv 1 \quad \text{iff n is the first integer such that } U_n(g_n) \leqslant \varepsilon \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{for all } (\underline{x}^n, \underline{e}^n) \\ &\equiv 0 \quad \text{otherwise} \end{aligned} \tag{4.16}$$

$n = 0,1,\ldots,$ is well-defined. Then the class of procedures $\{(\underline{\phi}^\varepsilon, \underline{\gamma}', \underline{\delta}^*) : \underline{\gamma}' \in \Gamma', \varepsilon > 0\}$ is essentially complete for the problem in which the cost is waived, for, given any procedure $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$, take $\varepsilon = r(G, (\underline{\phi}, \underline{\gamma}, \underline{\delta}))$. Then clearly

$$r(G, (\underline{\phi}^\varepsilon, \underline{\gamma}', \underline{\delta}^*)) \leqslant r(G, (\underline{\phi}, \underline{\gamma}, \underline{\delta})) \quad \text{for any } \underline{\gamma}' \in \Gamma' . \tag{4.17}$$

Therefore, the goal of the statistician is modified as follows; with the cost set to zero, it is desired to find a procedure $(\underline{\phi}, \underline{\gamma}, \underline{\delta})$ for which

$$r(G, (\underline{\phi}, \underline{\gamma}, \underline{\delta})) \leqslant \varepsilon \tag{4.18}$$

where $\varepsilon > 0$ is some prespecified constant. Equivalently (since $\underline{\delta}^*$ will obviously be the terminal decision rule used) it is desired to find a procedure $(\underline{\phi}, \underline{\gamma})$ for which

$$U_N(g_N) \leqslant \alpha \tag{4.19}$$

where α, $0 < \alpha < 1$, is some prespecified constant. (Note that if $\alpha \geqslant 1$, then it follows from Corollary 4.1 that (4.19) is satisfied by any $\underline{\gamma}$ and the stopping rule which stops before any observations are taken.) Using (4.19) as the requirement of the modified goal, it follows that the stopping rule $\underline{\phi}^{\alpha}$ defined by (4.16) (with the stopping requirement no longer restricted to all $(\underline{x}^n, \underline{e}^n)$) will attain the modified goal, provided that the experiment rule $\underline{\gamma}$ used is a member of Γ'. It is clear that $\underline{\phi}^{\alpha}$ does not depend on backward induction methods for its solution and is easily computed.

From the above, it follows that any experiment rule $\underline{\gamma} \in \Gamma'$ will satisfy the goal (4.19) if used in conjunction with the stopping rule $\underline{\phi}^{\alpha}$ and the terminal decision rule $\underline{\delta}^*$. The optimality criterion used for choosing amongst experiment rules is that of minimising the expected sample size; thus, $(\underline{\phi}^*, \underline{\gamma}^*, \underline{\delta}^*)$ is said to be optimal for the modified goal if

$$E_{G, f_\theta, \underline{\phi}^{\alpha}, \underline{\gamma}^*}[N] \leqslant E_{G, f_\theta, \underline{\phi}^{\alpha}, \underline{\gamma}}[N] \tag{4.20}$$

for all experiment rules $\underline{\gamma}$ for which $(\underline{\phi}^{\alpha}, \underline{\gamma}, \underline{\delta}^*)$ satisfies the modified goal (4.19). (It may be taken that Γ' is the class of all such experiment rules).

Interpretation Of The Modified Goal

From (4.19) and (4.20), it follows that the modified goal of the statistician is to ensure that the expected posterior loss (i.e. the posterior probability of error) is no greater than some prespecified value at termination, and, subject to this constraint, to cease experimentation as soon as possible. In effect, (4.20) "compensates" for the elimination of the sampling cost.

It might be claimed that modifying the goal as above merely "manoeuvres" the mathematical formulation from the usual decision-theoretic model of expected loss-plus-cost to that of (many-) hypothesis testing. However, it is clear from the Normal example considered in Chapter 3 that global minimisation of the risk is an impossible task computationally. Faced with this situation, the statistician might well settle for an "acceptable" risk and a "small" sample size. Furthermore, by eliminating the need to solve backward induction equations in order to compute the optimal stopping rule, it is possible to consider the experiment rule alone, and thus gain insight into what an optimal experimental design might be. This is because $\underline{\phi}^{\alpha}$ is independent of the experiment rule $\underline{\gamma}$ in the sense that, as sampling progresses, the evaluation of the stopping rule at each stage does not depend on $\underline{\gamma}$, only on the current expected loss $U_n(g_n)$. Therefore, it is now possible to consider the performance of the OSA experiment rule $\underline{\gamma}^0$.

3. CHARACTERISATION OF THE CLASS Γ'

In order to investigate the performance of experiment rules which belong to the class Γ', and in particular, to ensure that the OSA rule $\underline{\gamma}^0$ is in fact a member of Γ', it is necessary to identify Γ' in some way. It is shown below that if Θ is finite and the population densities $f_\theta(x|e)$ are "different" for different values of θ (in the sense of (4.21) below), then any experiment rule $\underline{\gamma}$ ensures the convergence of $U_n(g_n)$ to zero as $n \to \infty$, i.e. Γ' coincides with Γ, the space of all experiment rules. In this case, $\underline{\gamma}^0$ is clearly a member of Γ'.

In what follows, θ_0 denotes the true value of the state of nature θ, and Θ_0 denotes the member of the partition $\{\Theta_1, \Theta_2, \ldots, \Theta_k\}$ of Θ for which $\theta_0 \in \Theta_0$.

Theorem 4.3 For any experiment rule $\underline{\gamma}$, in order that $U_n(g_n) \to 0$ as $n \to \infty$ under $\underline{\gamma}$, it is sufficient that $P(\theta_0|\underline{x}^n, \underline{e}^n) \to 1$ as $n \to \infty$ under $\underline{\gamma}$.

Proof: This follows immediately from the Corollary to Theorem 4.1. □

The result that $\Gamma' = \Gamma$, given in Theorem 4.4 below, is shown in the notation of Chernoff (1959) and Bessler (1960), which was introduced in Chapter 1. Both proved similar results in these papers.

Theorem 4.4 If the two conditions

a) For every $\theta_i, \theta_j \in \Theta$ such that $\theta_i \neq \theta_j$ and every $e \in \mathcal{E}$,

$$I(\theta_i, \theta_j, e) = \int \log\left[\frac{f_{\theta_i}(x|e)}{f_{\theta_j}(x|e)}\right] f_{\theta_i}(x|e)\,dx > 0 \tag{4.21}$$

b) There exists an M such that, for any $\theta_i(\neq\theta_0) \in \Theta$ and any $e \in \mathcal{E}$,

$$E\left[\left\{\log \frac{f_{\theta_0}(X|e)}{f_{\theta_i}(X|e)}\right\}^2\right] < M \quad \text{where } E \equiv E_{\theta_0} \tag{4.22}$$

hold, then $g(\theta_0|\underline{x}^n, \underline{e}^n) \to 1$ as $n \to \infty$ for any experiment rule $\underline{\gamma}$.

Proof: By definition of the posterior probability density,

$$g(\theta_0|\underline{x}^n, \underline{e}^n) = \frac{g_0 \prod_{j=1}^{n} f_{\theta_0}(x_j|e_j)}{\sum_{r=1}^{k}\left[g_r \prod_{j=1}^{n} f_{\theta_r}(x_j|e_j)\right]}$$

$$= \left\{1 + \sum_{\substack{r=1 \\ "r\neq 0"}}^{k} g_r \prod_{j=1}^{n} \frac{f_{\theta_r}(x_j|e_j)}{f_{\theta_0}(x_j|e_j)}\right\}^{-1} \tag{4.23}$$

where it is assumed without loss of generality that $\Theta = \{\theta_1, \ldots, \theta_k\}$.

Hence it suffices to show that

$$\lim_{n\to\infty} \prod_{j=1}^{n} \frac{f_{\theta_r}(x_j|e_j)}{f_{\theta_0}(x_j|e_j)} = 0 \quad \text{for every } \theta_r(\neq\theta_0) \in \Theta. \tag{4.24}$$

To show this, note that Condition a) implies that

$$\mu_r(e) = E_{\theta_0}\left\{\log \frac{f_{\theta_r}(x|e)}{f_{\theta_0}(x|e)}\right\} < 0 \qquad \text{for all } \theta_r(\neq\theta_0) \in \Theta \text{ and all } e \in \mathcal{E} \tag{4.25}$$

and, in conjunction with Condition b), this implies that

$$\mathrm{Var}_{\theta_0}\left\{\log \frac{f_{\theta_r}(x|e)}{f_{\theta_0}(x|e)}\right\} < M - \mu_r(e)^2 \qquad \text{for all } \theta_r(\neq\theta_0) \in \Theta \text{ and all } e \in \mathcal{E}. \tag{4.26}$$

Thus, for any $\theta_r(\neq\theta_0) \in \Theta$, it follows from a theorem of Kolmogorov (see e.g. Fisz (1963), Theorem 6.12.1) that

$$\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{j=1}^{n}\log\frac{f_{\theta_r}(x_j|e_j)}{f_{\theta_0}(x_j|e_j)} - k_n\right\} = 0 \qquad (\text{a.s. } F) \tag{4.27}$$

where

$$\begin{aligned} k_n &= \frac{1}{n}\sum_{j=1}^{n} E_{\theta_0}\left\{\log\frac{f_{\theta_r}(X_j|e_j)}{f_{\theta_0}(X_j|e_j)}\right\} \\ &= \frac{1}{n}\sum_{j=1}^{n}\mu_r(e_j) \\ &< 0 \ . \qquad \text{from (4.25)} \end{aligned} \tag{4.28}$$

Therefore

$$\begin{aligned} \log \lim_{n\to\infty}\prod_{j=1}^{n}\frac{f_{\theta_r}(x_j|e_j)}{f_{\theta_0}(x_j|e_j)} &= \lim_{n\to\infty}\log\prod_{j=1}^{n}\frac{f_{\theta_r}(x_j|e_j)}{f_{\theta_0}(x_j|e_j)} \\ &= -\infty \qquad (\text{a.s. } F) \end{aligned} \tag{4.29}$$

from which the desired result (4.24) follows. □

Remarks: Theorem 4.4 is essentially the generalisation to the design case of the well-known consistency property of the maximum-likelihood estimator $\hat{\theta}_n$ of θ_0. It may be shown that this consistency is exponential; see Chernoff (1959), Lemma 1. In order to ensure consistency in the design case for any experiment rule γ, however, Condition a) must hold; for if

there exists an experiment $e \in \mathcal{E}$ and a $\theta_r (\neq \theta_0) \in \Theta$ such that $I(\theta_0, \theta_r, e)$ is zero, then the experiment rule γ_e, which selects experiment e at every stage with probability one, will not be a member of Γ', as no amount of experimentation will be sufficient for the experimenter to distinguish between θ_r and θ_0.

4. IDENTIFYING ONE OF TWO HYPOTHESES CONCERNING THE PARAMETERS OF TWO BINOMIAL POPULATIONS

(1) Set-up And Notation

This is the two-population analogue to the identification problem described in Chapter 3. There are 2 populations Π_1 and Π_2, with the associated binomial parameters θ_1 and θ_2 respectively. The state space $\Theta = \{\underline{\theta}_1, \underline{\theta}_2\} = \{(\nu_2, \nu_1), (\nu_1, \nu_2)\}$, where ν_1 and ν_2 $(\nu_2 > \nu_1)$ are two known values in the open interval (0,1). Then $\underline{\theta} = \underline{\theta}_i$ iff $\theta_i = \nu_2$, $i = 1,2$, and the partition $\Theta = \{\Theta_1, \Theta_2\}$ is again defined by $\Theta_i = \{\underline{\theta}_i\}$, $i = 1,2$. The known prior probability that $\underline{\theta}$ equals $\underline{\theta}_1$ is denoted by g, $0 < g < 1$.

Once again, $X_1, X_2, \ldots, X_m, \ldots$ denotes a sequence of i.i.d. observations from Π_1 and $Y_1, Y_2, \ldots, Y_n, \ldots$ denotes a sequence of i.i.d observations from Π_2 with respective probability functions

$$f(x|\theta_1) = \theta_1^x (1-\theta_1)^{1-x}, \quad x = 0,1$$
$$f(y|\theta_2) = \theta_2^y (1-\theta_2)^{1-y}, \quad y = 0,1. \tag{4.30}$$

Furthermore, the sequence $\{T_n\}$ defined for each n by $t_n = \{s_{1,n}, f_{1,n}, s_{2,n}, f_{2,n}\}$ is PARS and POLS for the sequential design problem (see (3.5)), and the posterior probability of $\underline{\theta}_1$ given n observations is (cf. (3.8) and (3.9))

$$g(\underline{\theta}_1 | t_n) = \frac{g\nu_2^{s_{1,n}}\mu_2^{f_{1,n}}\nu_1^{s_{2,n}}\mu_1^{f_{2,n}}}{\left[g\nu_2^{s_{1,n}}\mu_2^{f_{1,n}}\nu_1^{s_{2,n}}\mu_2^{f_{2,n}} + (1-g)\nu_2^{s_{2,n}}\mu_2^{f_{2,n}}\nu_1^{s_{1,n}}\mu_1^{f_{1,n}}\right]}$$

$$= \frac{F_{1,n}}{F_{1,n} + F_{2,n}} \tag{4.31}$$

where $\mu_i = 1 - \nu_i$, $i = 1,2$.

To show that Conditions a) and b) of Theorem 4.4 are satisfied, note that for $(i,j) = (1,2)$ or $(2,1)$,

$$I(\underline{\theta}_i, \underline{\theta}_j, e) = \begin{cases} (1-\nu_2)\log\left(\frac{1-\nu_2}{1-\nu_1}\right) + \nu_2\log\left(\frac{\nu_2}{\nu_1}\right) & \text{if } e = i \\ (1-\nu_1)\log\left(\frac{1-\nu_1}{1-\nu_2}\right) + \nu_1\log\left(\frac{\nu_1}{\nu_2}\right) & \text{if } e = j \end{cases} \tag{4.32}$$

(cf. Bessler (1960) part 2, p.14). Therefore, Condition a) holds if

$$(1-\nu_2)^{1-\nu_2}\nu_2^{\nu_2} > (1-\nu_1)^{1-\nu_2}\nu_1^{\nu_2}$$

and (4.33)

$$(1-\nu_1)^{1-\nu_1}\nu_1^{\nu_1} > (1-\nu_2)^{1-\nu_1}\nu_2^{\nu_1} .$$

For any a, $0 < a < 1$, define the function $f_a(x)$ on $(0,1)$ by

$$f_a(x) = x^a(1-x)^{1-a} \tag{4.34}$$

which is positive-valued and differentiable on $(0,1)$. Further,

$$f_a'(x) = ax^{a-1}(1-x)^{1-a} - (1-a)x^a(1-x)^{-a}$$

$$= 0 \tag{4.35}$$

has the unique solution $x = a$ on $(0,1)$. Since $f_a(0) = f_a(1) = 0$, it follows that $f_a(x)$ has its maximum value at $x = a$, for any a, $0 < a < 1$. Thus (4.33) follows and assumption a) is satisfied. Condition b) is shown to hold in this case in Bessler (1960) part 2, p.19; from (4.32), it is clear that

$$M = (1-\nu_2)\log\left(\frac{1-\nu_2}{1-\nu_1}\right)^2 + \nu_2\log\left(\frac{\nu_2}{\nu_1}\right)^2$$

$$+ (1-\nu_1)\log\left(\frac{1-\nu_1}{1-\nu_2}\right)^2 + \nu_1\log\left(\frac{\nu_1}{\nu_2}\right)^2 \qquad (4.36)$$

will satisfy (4.22). Therefore, any experiment rule for this identification problem belongs to Γ', and in particular, $\underline{\gamma}^0$ is a member of Γ'.

(2) An Optimality Result Concerning $\underline{\gamma}^0$

The One-Step-Ahead Experiment Rule $\underline{\gamma}^0$. The OSA experiment rule was defined in (4.1); however, it follows from Theorem 4.1 that $\underline{\gamma}^0$ may be described as follows:

"after each stage of experimentation n, n = 1,2,..., perform experiment i next if

$$E\left[\max_{1\leqslant j\leqslant k}\left\{P(\theta_j|\underline{x}^n, X, \underline{e}^n, e_{n+1} = i)\right\}|\underline{x}^n, \underline{e}^n\right]$$

$$= \max_{1\leqslant r\leqslant k} E\left[\max_{1\leqslant j\leqslant k}\left\{P(\theta_j|\underline{x}^n, X, \underline{e}^n, e_{n+1} = r)\right\}|\underline{x}^n, \underline{e}^n\right],$$

$$i = 1,2,..., k". \qquad (4.37)$$

In the two-population case, this may be written as

"perform experiment 1 next if

$$E\left[\max_{i=1,2}\left\{g(\theta_i|\underline{x}^m, X, \underline{y}^n)\right\}|\underline{x}^m, \underline{y}^n\right]$$

$$\geqslant E\left[\max_{i=1,2}\left\{g(\theta_i|\underline{x}^m, \underline{y}^n, Y)\right\}|\underline{x}^m, \underline{y}^n\right]$$

otherwise perform experiment 2." * (4.38)

Viewed in this fashion, it may be seen that the OSA rule selects next the experiment giving rise to the larger expected value of the maximum posterior probability at the next stage. Therefore, one might conjecture

*If the two expectations are equal, of course, either experiment may be performed.

that this procedure maximises the probability that sampling is terminated (given that $(\underline{X}^n, \underline{E}^n) = (\underline{x}^n, \underline{e}^n)$) at the next (i.e. $(n+1)$th) stage. The following theorem shows that this is true for the two-binomial identification problem for a large range of values of ν_1 and ν_2.

Theorem 4.5 For certain values of ν_1, ν_2 and t in $(0,1)^2$ (defined in (4.66) below), the probabilities $P(\max_{i=1,2}\{g_i(X)\} \geqslant t)$ and $P(\max_{i=1,2}\{g_i(Y)\} \geqslant t)$ are non-decreasing functions of $E[\max_{i=1,2}\{g_i(X)\}]$ and $E[\max_{i=1,2}\{g_i(Y)\}]$ respectively (where $g_i(\cdot)$ is a convenient notation for $g(\underline{\theta}_i|\cdot)$, $i = 1,2$).

Remark: As is well-known, the posterior probabilities $g(\underline{\theta}_i|\underline{x}^m, \underline{y}^n)$ may be thought of as prior probabilities $\hat{g}(\underline{\theta}_i)$ in the sense that the problem for the experimenter who is at the (m,n)th stage and has observed the r.v.s $(\underline{X}^m, \underline{Y}^n) = (\underline{x}^m, \underline{y}^n)$, and thus the posterior probabilities $g(\underline{\theta}_i|\underline{x}^m, \underline{y}^n)$, is equivalent to the problem for the experimenter at the "zero-th" stage, and whose prior probabilities are $\hat{g}(\underline{\theta}_i) = g(\theta_i|\underline{x}^m, \underline{y}^n)$, $i = 1,2$.

Proof: Without loss of generality the result is proved for the "X" observation; the proof for a "Y" observation follows from the symmetry of the problem. Defining $\nu = \nu_1/\nu_2$ and $\mu = \mu_1/\mu_2$, and noting that $\nu < 1$ and $\mu > 1$, the posterior probabilities $g_i(X)$ may be written as

$$g_1(X) = \begin{cases} \left(1 + \mu \dfrac{1-g}{g}\right)^{-1} & \text{if } X = 0 \\ \left(1 + \nu \dfrac{1-g}{g}\right)^{-1} & \text{if } X = 1 \end{cases}$$

$$g_2(X) = \begin{cases} \left(1 + \dfrac{1}{\mu} \dfrac{g}{1-g}\right)^{-1} & \text{if } X = 0 \\ \left(1 + \dfrac{1}{\nu} \dfrac{g}{1-g}\right)^{-1} & \text{if } X = 1 \end{cases} \qquad (4.39)$$

from which it follows that

$$g_1(0) < g_1(1) \quad \text{and} \quad g_2(0) > g_2(1). \tag{4.40}$$

There are three cases to consider:

a) $g > \dfrac{\mu_1}{\mu_1 + \mu_2} = \dfrac{\mu}{1+\mu}$: (4.41)

This implies that

$$g_1(0) > g_2(0)$$

and hence

$$g_2(1) < g_2(0) < g_1(0) < g_1(1). \tag{4.42}$$

Therefore, in this case

$$\begin{aligned} E\left[\max_{i=1,2} g_i(X)\right] &= g_1(0)P(X=0) + g_1(1)P(X=1) \\ &= g \end{aligned} \tag{4.43}$$

where $P(X=0) = gf(0|\nu_2) + (1-g)f(0|\nu_1)$ is the probability that $X=0$ unconditional on $\underline{\theta}$, with a similar expression for $P(X=1)$. Consequently, it must be shown that $P(\max_{i=1,2}\{g_i(X)\} \geqslant t)$ is a non-decreasing function of g. Firstly

$$P\left(\max_{i=1,2}\{g_i(X)\} \geqslant t\right) = 0$$

$$\text{iff} \quad g_1(1) < t \qquad \text{(from (4.42))}$$

$$\text{iff} \quad 1 + \nu\,\frac{1-g}{g} > \frac{1}{t} \qquad \text{(from (4.39))}$$

$$\text{iff} \quad g < \frac{\nu t}{1 - t + \nu t} \tag{4.44}$$

Similarly

$$P\left(\max_{i=1,2}\{g_i(X)\} \geqslant t\right) = 1$$

$$\text{iff} \quad g_1(0) \geqslant t$$

$$\text{iff} \quad g \geqslant \frac{\mu t}{1 - t + \mu t} \, . \tag{4.45}$$

Finally

$$P(\max_{i=1,2}\{g_i(X)\} \geqslant t) = P(X=1)$$

$$\text{iff} \quad g_1(0) < t \leqslant g_1(1)$$

$$\text{iff} \quad \frac{\nu t}{1 - t + \nu t} \leqslant g < \frac{\mu t}{1 - t + \mu t} \, . \tag{4.46}$$

Furthermore, it can be shown that

$$\nu_2 > \nu_1 \quad \text{implies} \quad \frac{\nu t}{1 - t + \nu t} < \frac{\mu t}{1 - t + \mu t} \quad \text{for any } t,\ 0 < t < 1. \tag{4.47}$$

Therefore, as these are the only possibilities, it follows that $P(\max_{i=1,2}\{g_i(X)\} \geqslant t)$ is a non-decreasing function of $E[\max_{i=1,2}\{g_i(X)\}]$ in this case.

b) $g < \dfrac{\nu_1}{\nu_1 + \nu_2} = \dfrac{\nu}{1+\nu}$: (4.48)

The proof in this case follows similarly to the proof in case a). The posterior probabilities satisfy

$$g_1(0) < g_1(1) < g_2(1) < g_2(0) \tag{4.49}$$

and hence

$$E[\max_{i=1,2}\{g_i(X)\}] = g_2(0)P(X=0) + g_2(1)P(X=1)$$
$$= 1 - g \tag{4.50}$$

and it must be shown that $P(\max_{i=1,2}\{g_i(X)\} \geqslant t)$ is a non-decreasing function of $1 - g$. Firstly

$$P(\max_{i=1,2}\{g_i(X)\} \geqslant t) = 0$$

$$\text{iff} \quad g_2(0) < t$$

$$\text{iff} \quad 1-g < \frac{\mu^{-1}t}{1-t+\mu^{-1}t} . \tag{4.51}$$

Secondly

$$P(\max_{i=1,2}\{g_i(X) \geqslant t\}) = 1$$

$$\text{iff} \quad g_2(1) \geqslant t$$

$$\text{iff} \quad 1-g \geqslant \frac{\nu^{-1}t}{1-t+\nu^{-1}t} . \tag{4.52}$$

Finally

$$P(\max_{i=1,2}\{g_i(X)\} \geqslant t) = P(X=0)$$

$$\text{iff} \quad g_2(1) < t \leqslant g_2(0)$$

$$\text{iff} \quad \frac{\mu^{-1}t}{1-t+\mu^{-1}t} \leqslant 1-g < \frac{\nu^{-1}t}{1-t+\nu^{-1}t} . \tag{4.53}$$

Again it may be shown that

$$\nu_2 > \nu_1 \quad \text{implies} \quad \frac{\mu^{-1}t}{1-t+\mu^{-1}t} < \frac{\nu^{-1}t}{1-t+\nu^{-1}t} \quad \text{for any } t,\ 0 < t < 1 \tag{4.54}$$

and as these are the only possibilities, $P(\max_{i=1,2}\{g_i(X)\} \geqslant t)$ is thus a non-decreasing function of $E[\max_{i=1,2}\{g_i(X)\}]$ in this case.

c) $\dfrac{\nu_1}{\nu_1+\nu_2} < g < \dfrac{\mu_1}{\mu_1+\mu_2}$: (4.55)

In this case

$$g_2(0) > \max\{g_1(0),\ g_2(1)\} \quad \text{and} \quad g_1(1) > \max\{g_1(0),\ g_2(1)\} \tag{4.56}$$

but there is no total ordering between the four function values. It follows that

$$\begin{aligned} E[\max_{i=1,2}\{g_i(X)\}] &= g_2(0)P(X=0) + g_1(1).P(X=1) \\ &= \mu_1(1-g) + \nu_2 g . \end{aligned} \tag{4.57}$$

If it is assumed without loss of generality that $\nu_2 > \mu_1$, then (4.57) may be written as

$$E[\max_{i=1,2}\{g_i(X)\}] = (\nu_2 - \mu_1)g + \mu_1. \tag{4.58}$$

Thus, it must be shown that $P(\max_{i=1,2}\{g_i(X)\} \geqslant t)$ is a non-decreasing function of g. Firstly

$$P(\max_{i=1,2}\{g_i(X)\} \geqslant t) = 0$$

$$\text{iff} \quad \max\{g_2(0), g_1(1)\} < t$$

$$\text{iff} \quad \frac{1-t}{1-t+\mu^{-1}t} < g < \frac{\nu t}{1-t+\nu t}. \tag{4.59}$$

Secondly

$$P(\max_{i=1,2}\{g_i(X)\} \geqslant t) = 1$$

$$\text{iff} \quad \min\{g_2(0), g_1(1)\} \geqslant t$$

$$\text{iff} \quad g \leqslant \frac{1-t}{1-t+\mu^{-1}t} \quad \underline{\text{or}} \quad g \geqslant \frac{\nu t}{1-t+\nu t}. \tag{4.60}$$

Thirdly

$$P(\max_{i=1,2}\{g_i(X)\} \geqslant t) = P(X=0)$$

$$\text{iff} \quad g_1(1) < t \leqslant g_2(0)$$

$$\text{iff} \quad g < \frac{\nu t}{1-t+\nu t} \quad \underline{\text{and}} \quad g \leqslant \frac{1-t}{1-t+\mu^{-1}t}. \tag{4.61}$$

Finally

$$P(\max_{i=1,2}\{g_i(X)\} \geqslant t) = P(X=1)$$

$$\text{iff} \quad g_2(0) < t \leqslant g_1(1)$$

$$\text{iff} \quad g > \frac{1-t}{1-t+\mu^{-1}t} \quad \underline{\text{and}} \quad g \geqslant \frac{\nu t}{1-t+\nu t}. \tag{4.62}$$

Therefore, it does not appear that the desired result holds in case c). However, note that $P(\max_{i=1,2}\{g_i(X)\} \geqslant t)$ is equal to 1 for any value of $t \leqslant .5$, since at least one of the posterior probabilities $g_1(x)$ and $g_2(x) = 1 - g_1(x)$ must be no less than .5. This may be proved formally by showing that

$$F(t) = \frac{\nu t}{1 - t + \nu t} \quad \text{and} \quad G(t) = \frac{1 - t}{1 - t + \mu^{-1} t} \tag{4.63}$$

are increasing and, respectively, decreasing functions of t. Since $F(.5) = \frac{\nu_1}{\nu_1 + \nu_2}$ and $G(.5) = \frac{\mu_1}{\mu_1 + \mu_2}$, $F(0) = 0$ and $G(0) = 1$, it follows that (4.60) always holds in case c) for $t \leqslant .5$.

Consequently, since case c) represents the situation in which the current (prior or posterior) probability of both $\underline{\theta}_1$ and $\underline{\theta}_2$ is near .5, it seems reasonable to consider those values of ν_1, ν_2 and t (where $t = 1 - \alpha$ is close to 1) which satisfy

$$P(\max_{i=1,2}\{g_i(X)\} \geqslant t) = 0 \quad \text{for all } g \text{ satisfying (4.55).} \tag{4.64}$$

From (4.55) and (4.59), this is true iff

$$\frac{1 - t}{1 - t + \mu^{-1} t} < \frac{\nu_1}{\nu_1 + \nu_2} \quad \underline{\text{and}} \quad \frac{\nu t}{1 - t + \nu t} > \frac{\mu_1}{\mu_1 + \mu_2}. \tag{4.65}$$

Both these inequalities hold (i.e. they have the same solution) iff

$$t > \frac{1}{1 + \nu\mu^{-1}} = \left(1 + \frac{\nu_1 \mu_2}{\nu_2 \mu_1}\right)^{-1}. \tag{4.66}$$

For the typical value of $t = .95$, the set of (ν_1, ν_2) satisfying (4.66) is

$$S = \{(\nu_1, \nu_2) : 0 < \nu_1 < \nu_2 < 1 \quad \text{and} \quad \nu_2 + 18\nu_1\nu_2 - 19\nu_1 < 0\} \tag{4.67}$$

which is shown in Figure 4.1 below.

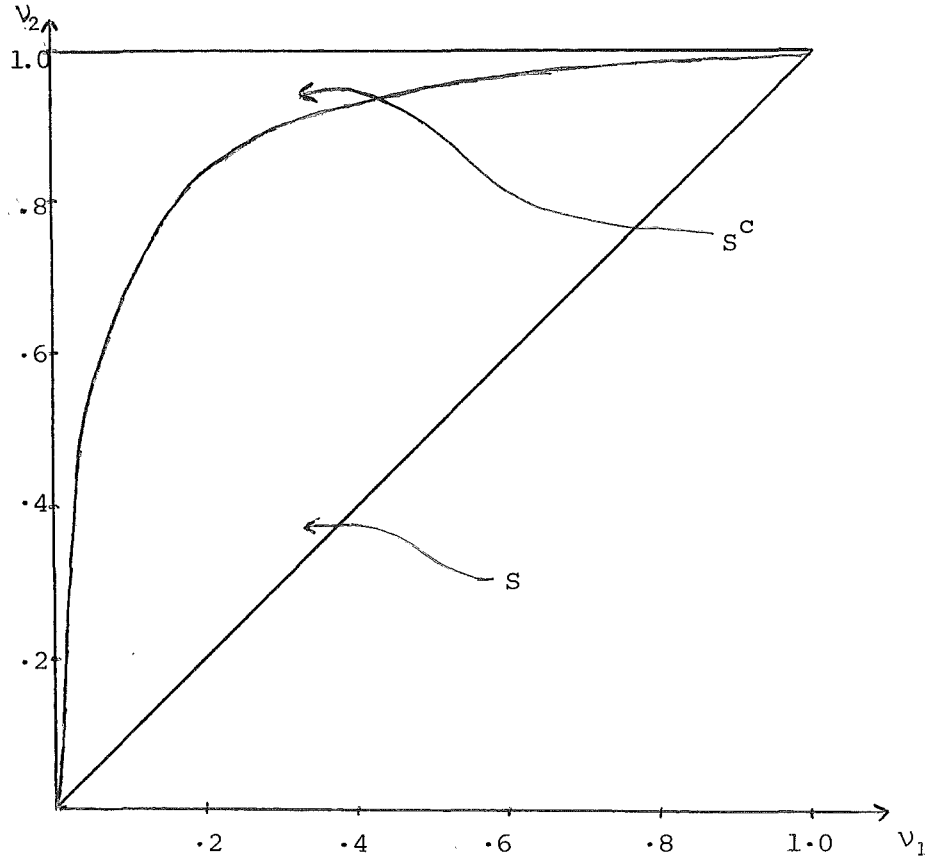

FIGURE 4.1. The region S defined in (4.67).

Therefore, if $(\nu_1, \nu_2) \in S$, it follows in case c) that $P(\max_{i=1,2}\{g_i(X)\} \geq t) = 0$ for all t, $.95 \leq t < 1$, which is non-decreasing in $E[\max_{i=1,2}\{g_i(X)\}]$. Furthermore, the same solution (4.66) is obtained if it is assumed that $\nu_2 < \mu_1$ and a similar analysis is carried out in terms of $1 - g$. □

Define

$$S_n = \{(\underline{x}^n, \underline{e}^n) : \max_{i=1,2}\{g(\theta_i | \underline{x}^n, \underline{e}^n)\} \geq P^*\}, \quad n = 0,1,... \tag{4.68}$$

to be the subset of $\mathcal{X}^n \times \mathcal{E}^n$ for which the modified goal is attained ($P^* = 1 - \alpha$ is the desired posterior "level of confidence", $0 < P^* < 1$). It follows that the stopping time N satisfies

$$\{N = n\} = \left\{S_n \cap \left(\bigcap_{i=0}^{n-1} S_i^c\right)\right\}, \quad n = 1,2,... \tag{4.69}$$

and that Theorem 4.5 implies that (if ν_1, ν_2 satisfy (4.66) with $t = P^*$)

$$P_{G,f_\theta,\underline{\phi}^\alpha,\underline{\gamma}^0}(S_{n+1}|F_n) \geqslant P_{G,f_\theta,\underline{\phi}^\alpha,\underline{\gamma}}(S_{n+1}|F_n) \quad \text{for any experiment rule } \underline{\gamma}, \tag{4.70}$$

$n = 0,1,\ldots$. It does not follow that $\underline{\gamma}^0$ is optimal in the sense of (4.20); however, the optimality property (4.70) would seem to be desirable, and it suggests that $\underline{\gamma}^0$ might give rise to a smaller expected sample size than that of many other experiment rules. For example, (4.70) is equivalent to the result that $\underline{\gamma}^0$ minimises the expected distance of the process from the stopping boundaries at the next stage of experimentation. Note from (4.31) that

$$g(\underline{\theta}_1|t_n) \geqslant P^*$$

$$\text{iff} \quad \left(\frac{\mu_1}{\mu_2}\right)^{f_2-f_1}\left(\frac{\nu_2}{\nu_1}\right)^{s_2-s_1} \geqslant \frac{P^*(1-g)}{g(1-P^*)}$$

$$\text{iff} \quad (f_2-f_1)\log\mu + (s_2-s_1)\log\nu \geqslant \log\left(\frac{1-g}{g}\right) + \log\left(\frac{P^*}{1-P^*}\right) \tag{4.71}$$

where the subscript n is omitted for notational convenience. Defining

$$\hat{x} = f_1 - f_2 \qquad \hat{y} = s_1 - s_2$$

$$a = \log\mu \qquad b = -\log\nu \qquad (a,b > 0)$$

$$c_1 = \log\left(\frac{1-g}{g}\right) \qquad c_2 = \log\left(\frac{P^*}{1-P^*}\right) > 0 \tag{4.72}$$

where $-c_2 < c_1 < c_2$ for a non-trivial problem, it follows that

$$g(\underline{\theta}_1|t_n) = P^* \quad \text{iff} \quad -a\hat{x}+b\hat{y} = c_1+c_2, \text{ denoted } A_1 \tag{4.73}$$

Similarly

$$g(\underline{\theta}_2|t_n) = P^* \quad \text{iff} \quad -a\hat{x}+b\hat{y} = c_1-c_2, \text{ denoted } A_2 . \tag{4.74}$$

These stopping boundaries in $(\hat{x}, \hat{y})$-space are shown in Figure 4.2 below.

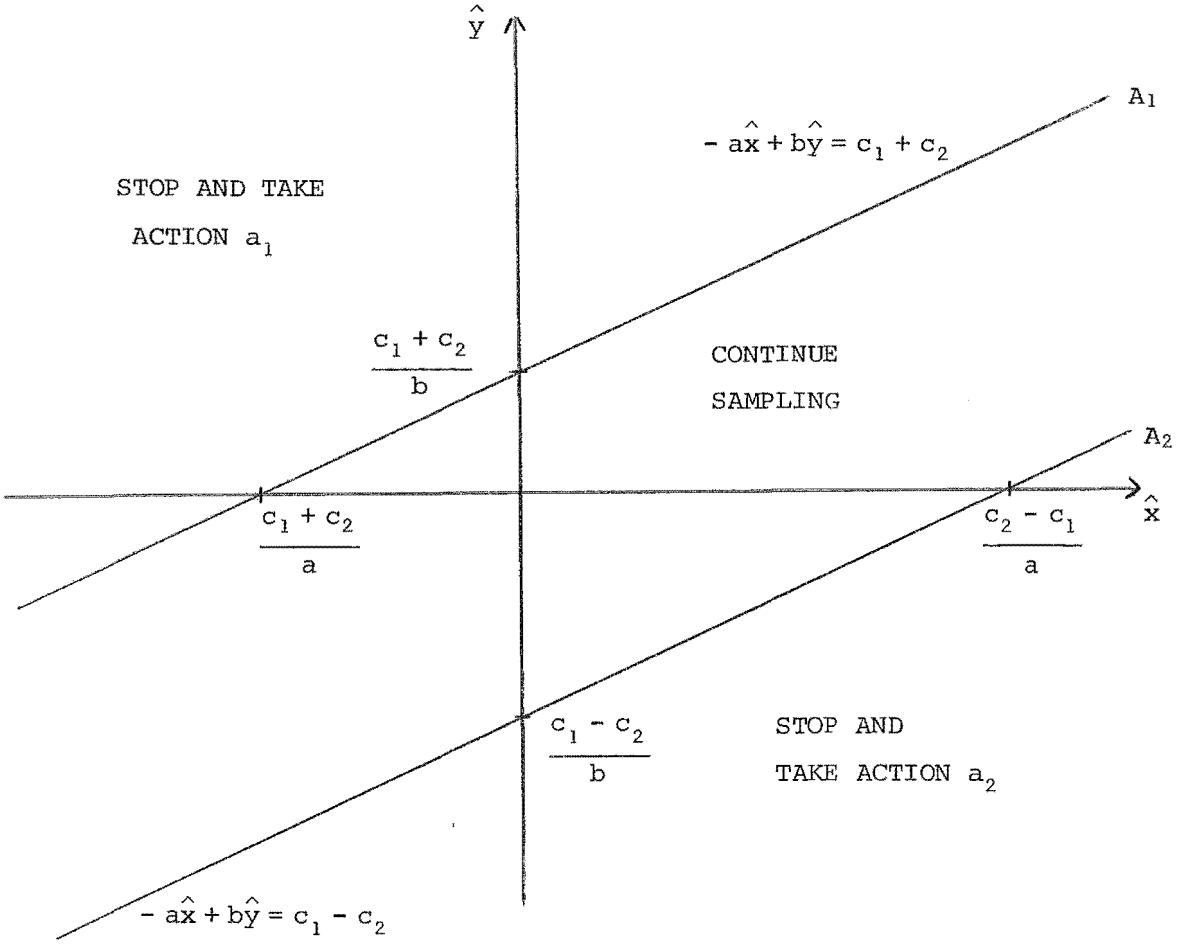

FIGURE 4.2. Continuation region and stopping regions in $(\hat{x},\hat{y})$-space.

Remarks: 1) Sampling from Π_1 means the process goes "up" the continuation region.

Sampling from Π_2 means the process goes "down" the continuation region.

2) $\nu_2 = 1-\nu_1$ iff $a = b$ iff the lines have gradient 1.

3) $g = .5$ iff $c_1 = 0$ iff x-intercepts are equal and y-intercepts are equal.

4) The state space is a grid $\{(\hat{x},\hat{y})$: $\hat{x}$ and $\hat{y}$ are integers$\}$; stopping occurs when a boundary is crossed.

Appropriate distances from a point $(\hat{x},\hat{y})$ in the continuation region to the

stopping boundaries are defined as follows:

$$D(\hat{y}, A_1) = \frac{c_1 + c_2 + a\hat{x}}{b} - \hat{y}$$

$$D(\hat{y}, A_2) = \hat{y} - \frac{c_1 - c_2 + a\hat{x}}{b}$$

$$D(\hat{x}, A_1) = \hat{x} - \frac{b\hat{y} - (c_1 + c_2)}{a}$$

$$D(\hat{x}, A_2) = \frac{b\hat{y} - (c_1 - c_2)}{a} - \hat{x} \ . \qquad (4.75)$$

Thus, if a further observation is taken, the $D(\hat{x}, .)$ distances change if a 0 is observed, and the $D(\hat{y}, .)$ distances change if a 1 is observed. Considering the population sample sizes m and n, it follows that the experiment rule which minimises the expected distance from the stopping boundaries at the next stage would choose experiment 1 if

$$E[\min_{i=1,2}\{D(X,A_i)\}] \leqslant E[\min_{i=1,2}\{D(Y,A_i)\}] \qquad (4.76)$$

where by an abuse of notation

$$D(X,A_i) = \begin{cases} D(\hat{x}+1,A_i) & \text{if } X = 0 \\ D(\hat{y}+1,A_i) & \text{if } X = 1 \end{cases} \quad , \quad i = 1,2 \qquad (4.77)$$

with a similar expression for $D(Y,A_i)$. Comparing "horizontal" distances, however, it follows that

$$D(\hat{x},A_1) < D(\hat{x},A_2)$$

$$\text{iff} \quad \frac{1-P^*}{P^*} \exp\{aD(\hat{x},A_1)\} < \frac{1-P^*}{P^*} \exp\{aD(\hat{x},A_2)\}$$

$$\text{iff} \quad g(\underline{\theta}_2|\cdot) < g(\underline{\theta}_1|\cdot) \qquad (4.78)$$

with the same result holding if "vertical" distances are compared (note that $aD(\hat{x},A_i) = bD(\hat{y},A_i)$, $i = 1,2$). Therefore, $\underline{\gamma}^0$ is the experiment rule which minimises the expected distances from the stopping boundaries at the next stage.

(3) Expressions For The Expected Sample Size

The following analysis, based on the work of Robbins (1974), shows that expressions for the expected sample size a) under pairwise sampling, and b) when $\nu_2 = 1 - \nu_1$, may be derived. From these results it may be shown that all experiment rules for this problem give rise to the same expected sample size when $\nu_2 = 1 - \nu_1$, as might be expected from the resulting symmetry of the problem.

Equal Sample Sizes. Suppose that at each stage of experimentation the number of observations taken from Π_1 equals the number of observations taken from Π_2; i.e. in the above notation, m = n. Then it is well-known that the Bayes rule (stopping and terminal decision) is equivalent to the following Sequential Probability Ratio Test (SPRT):

$$\text{take action } a_1 \text{ (say } \underline{\theta} = \underline{\theta}_1) \quad \text{if} \quad \left(\frac{\nu_2\mu_1}{\nu_1\mu_2}\right)^{s_{1,n}-s_{2,n}} \geqslant GR$$

$$\text{take action } a_2 \text{ (say } \underline{\theta} = \underline{\theta}_2) \quad \text{if} \quad \left(\frac{\nu_2\mu_1}{\nu_1\mu_2}\right)^{s_{1,n}-s_{2,n}} \leqslant GR^{-1}$$

$$\text{otherwise take a further observation from both } \Pi_1 \text{ and } \Pi_2^* \tag{4.79}$$

where $G = \frac{1-g}{g}$, $R = \frac{P^*}{1-P^*}$, and define $\lambda = \frac{\nu_2\mu_1}{\nu_1\mu_2} > 1$. Then

$$P(a_2 \text{ is taken} \,|\, \underline{\theta} = \underline{\theta}_1) = P\left(s_{1,N} - s_{2,N} = \frac{\log GR^{-1}}{\log \lambda} = -B \text{, say} \,|\, \underline{\theta}_1\right)$$

$$= \sum_{n=1}^{\infty} \sum_{\{N=n, s_{1,n}-s_{2,n}=-B\}} f(s_{1,n}, s_{2,n} | \underline{\theta}_1)$$

(assuming B is a positive integer)

$$= \sum_{n=1}^{\infty} \sum_{\{N=n, s_{1,n}-s_{2,n}=-B\}} \lambda^{s_{1,n}-s_{2,n}} f(s_{1,n}, s_{2,n} | \underline{\theta}_2)$$

$$= \lambda^{-B} P(a_2 \text{ is taken} \,|\, \underline{\theta} = \underline{\theta}_2) \tag{4.80}$$

*In this subsection, N and n refer to the number of pairs of observations

where

$$f(s_{1,n}, s_{2,n}|\underline{\theta}) = \theta_1^{s_{1,n}}(1-\theta_1)^{n-s_{1,n}}\theta_2^{s_{2,n}}(1-\theta_2)^{n-s_{2,n}} \tag{4.81}$$

is the joint density function of $s_{1,n}$ and $s_{2,n}$ (cf. (4.31)). Similarly

$$P(a_1 \text{ is taken} \mid \underline{\theta}=\underline{\theta}_2) = \lambda^{-C}P(a_1 \text{ is taken} \mid \underline{\theta}=\underline{\theta}_1) \tag{4.82}$$

where $C = \frac{\log GR}{\log \lambda}$ is assumed to be a positive integer. Therefore, the posterior probability of making an incorrect terminal decision using (4.79) is

$$gP(a_2 \text{ is taken} \mid \underline{\theta}_1) + (1-g)P(a_1 \text{ is taken} \mid \underline{\theta}_2)$$

$$= 1 - P^* \quad \text{iff}$$

$$g\lambda^{-B}P(a_2 \text{ is taken} \mid \underline{\theta}_2) + (1-g)P(a_1 \text{ is taken} \mid \underline{\theta}_2) = 1 - P^*$$

$$\text{iff} \quad P_2(1-g-g\lambda^{-B}) = 1 - P^* - \lambda^{-B}g$$

(where P_2 denotes $P(a_1 \text{ is taken} \mid \underline{\theta}_2)$)

$$\text{iff} \quad P_2 = \frac{[P^* - (1-g)](1-P^*)}{(2P^*-1)(1-g)}\,, \tag{4.83}$$

using the result

$$\lambda^{-B}g = g\frac{G}{R} = \frac{(1-g)(1-P^*)}{P^*}\,. \tag{4.84}$$

Similarly, it may be shown that

$$P_1 = P(a_2 \text{ is taken} \mid \underline{\theta}_1) = \frac{(P^*-g)(1-P^*)}{g(2P^*-1)}\,. \tag{4.85}$$

Now by Wald's lemma (Lemma 3.1, Chow, Robbins and Siegmund (1971))

$$\left.\begin{aligned} E[s_{1,N} - s_{2,N}|\underline{\theta}_1] &= (\nu_2 - \nu_1)E[N|\underline{\theta}_1] \\ E[s_{1,N} - s_{2,N}|\underline{\theta}_2] &= (\nu_1 - \nu_2)E[N|\underline{\theta}_2] \end{aligned}\right\} \tag{4.86}$$

and, ignoring overshoot of the stopping boundaries (i.e. assuming B and C are both positive integers)

$$\left.\begin{aligned} E[s_{1,N} - s_{2,N} | \underline{\theta}_1] &= C(1-P_1) - BP_1 \\ E[s_{1,N} - s_{2,N} | \underline{\theta}_2] &= -B(1-P_2) + CP_2 \end{aligned}\right\} . \tag{4.87}$$

Therefore, combining the last two pairs of equations,

$$\begin{aligned} E[N] &= gE[N|\underline{\theta}_1] + (1-g)E[N|\underline{\theta}_2] \\ &= \frac{1}{\nu_2 - \nu_1}\left\{g\left[C(1-P_1) - BP_1\right] + (1-g)\left[B(1-P_2) - CP_2\right]\right\} \\ &= \frac{B(P^*-g) + C[P^* - (1-g)]}{\nu_2 - \nu_1} \\ &= \frac{(2P^*-1)\log\left(\frac{P^*}{1-P^*}\right) + (2g-1)\log\left(\frac{1-g}{g}\right)}{(\nu_2 - \nu_1)\log\left(\frac{\nu_2\mu_1}{\nu_1\mu_2}\right)} \end{aligned} \tag{4.88}$$

The Design Case When $\nu_2 = 1 - \nu_1$

The test in this case is

take action a_1 if $(f_2 - f_1)\log\mu + (s_2 - s_1)\log\nu \geqslant \log GR$

take action a_2 if $(f_2 - f_1)\log\mu + (s_2 - s_1)\log\nu \leqslant \log GR^{-1}$

otherwise continue sampling . (4.89)

It may easily be shown that the two probabilities of error, P_1 and P_2, are the same as in the pairwise case, i.e. (4.83) and (4.85) still hold. Once again, Wald's Lemma gives

$$\left.\begin{aligned} E[(f_2 - f_1)\log\mu + (s_2 - s_1)\log\nu | \underline{\theta}_1] \\ = \log\mu[\mu_1 E_1 N - \mu_2 E_1 M] + \log\nu[\nu_1 E_1 N - \nu_2 E_1 M] \\ E[(f_2 - f_1)\log\mu + (s_2 - s_1)\log\nu | \underline{\theta}_2] \\ = \log\mu[\mu_2 E_2 N - \mu_1 E_2 M] + \log\nu[\nu_2 E_2 N - \nu_1 E_2 M] \end{aligned}\right\} \tag{4.90}$$

where $E_i M = E[M|\underline{\theta}_i]$, $i = 1,2$ etc., and ignoring overshoot of the stopping boundaries,

$$\left.\begin{aligned} E[(f_2 - f_1)\log\mu + (s_2 - s_1)\log\nu \mid \underline{\theta}_1] \\ = (1 - P_1)\log GR + P_1 \log GR^{-1} \\ E[(f_2 - f_1)\log\mu + (s_2 - s_1)\log\nu \mid \underline{\theta}_2] \\ = (1 - P_2)\log GR^{-1} + P_2 \log GR \end{aligned}\right\} . \tag{4.91}$$

Combining (4.90) and (4.91), and noting that $\mu_2 = \nu_1$, $\mu_1 = \nu_2$ and $\log\nu = -\log\mu$ here, yields

$$\left.\begin{aligned} (\nu_2 - \nu_1)\log\mu \cdot E_1[M+N] &= (1 - P_1)\log GR + P_1 \log GR^{-1} \\ (\nu_1 - \nu_2)\log\mu \cdot E_2[M+N] &= (1 - P_2)\log GR^{-1} + P_2 \log GR \end{aligned}\right\} . \tag{4.92}$$

Multiplying the first equation by g and the second by 1-g, it follows that

$$E[M+N] = \frac{(2P^* - 1)\log\left(\frac{P^*}{1-P^*}\right) + (2g-1)\log\left(\frac{1-g}{g}\right)}{(1-2\nu_1)\log\left(\frac{1-\nu_1}{\nu_1}\right)} \tag{4.93}$$

and this result holds for any experiment rule. Comparing (4.93) with (4.88), it follows that no experiment rule can do better (in terms of expected sample size) than pairwise sampling.

(4) The Performance Of $\underline{\gamma}^0$

Table 4.2 lists the performance of the OSA experiment rule for selected values of g, ν_1 and ν_2 under computer simulation. Also included are the corresponding values for the pairwise design rule (m = n) and the Bessler / Chernoff experiment rule defined by (1.4) and (1.5). The performance characteristics listed are the probability of correct decision or selection, usually denoted as P(CS), and the average sample size or number, usually denoted as ASN; these empirical values were obtained as sample means from the appropriate data values from 800 simulations in each case, with a similar calculation for the standard error of the ASN. Simulation was carried out on a Burroughs B6700 computer. The theoretical ASN was obtained by substituting the corresponding P(CS) value obtained

into the expression (4.88) (as P*); the resultant theoretical ASN value for a pairwise rule whose "P*-level" is the same as the rule in question may be used both to evaluate the efficiency of a design rule over a non-design pairwise rule (tabulated in the last column) and, in the case of the pairwise rule itself, to provide an estimate of error due to Monte

TABLE 4.2. Performance of OSA (first entry), Bessler's (second entry) and pairwise (third entry) experiment rules for certain values of g, ν_1, ν_2. $P^* = 0.95$. 800 simulations.

g	ν_1	ν_2	P(CS)	ASN	Theoretical ASN	Standard Error (ASN)	Theoretical ASN / ASN
.5	.2	.3	.958	100.2	105.8	2.56	1.06
			.954	107.1	101.9	2.65	.95
			.959	108.4	107.1	2.70	.99
.7	.2	.3	.953	89.0	88.1	2.64	.99
			.951	91.5	86.9	2.78	.95
			.949	91.9	84.6	2.61	.92
.9	.2	.3	.955	36.0	37.9	1.93	1.05
			.965	37.8	49.2	1.94	1.30
			.964	41.0	47.7	1.86	1.16
.5	.4	.55	.941	60.4	53.8	1.44	.89
			.944	60.3	55.1	1.51	.91
			.960	59.2	64.3	1.56	1.09
.7	.4	.55	.958	51.6	55.2	1.47	1.07
			.958	54.6	55.2	1.54	1.01
			.953	56.2	52.2	1.47	.93
.9	.4	.55	.954	22.7	21.7	1.10	.96
			.963	21.5	27.4	1.01	1.27
			.973	28.5	35.5	1.19	1.25

TABLE 4.2. continued

g	ν_1	ν_2	P(CS)	ASN	Theoretical ASN	Standard Error (ASN)	Theoretical ASN / ASN
.5	.5	.65	.964	59.8	68.3	1.46	1.14
			.964	62.3	65.5	1.62	1.05
			.949	57.3	56.4	1.42	.98
.7	.5	.65	.954	49.4	51.9	1.37	1.05
			.978	51.4	70.3	1.47	1.37
			.969	58.9	62.0	1.54	1.05
.9	.5	.65	.966	21.8	29.5	1.08	1.35
			.963	21.8	27.7	1.02	1.27
			.974	27.2	35.9	1.01	1.32
.5	.5	.9	.983	7.72	8.85	0.16	1.15
			.973	6.94	7.67	0.18	1.11
			.990	9.38	10.25	0.21	1.09
.7	.5	.9	.981	7.58	7.90	0.19	1.04
			.961	6.31	5.97	0.18	.95
			.959	5.70	5.80	0.16	1.02
.9	.5	.9	.981	4.57	4.67	0.17	1.02
			.981	4.68	4.67	0.16	1.00
			.983	5.50	4.85	0.18	0.87

Carlo simulation etc. This seems reasonable in the last case considered ($\nu_1 = .5$, $\nu_2 = .9$), as the "overshoot" from discrete data in a small sample size situation affects the corresponding P(CS) and ASN values to such a degree that comparison is impossible. Therefore, the "efficiency" of the procedure compared to pairwise sampling, defined as the ratio of the theoretical ASN to the observed ASN, is perhaps a better indicator of performance here.

From the results it may be seen that, in the cases considered, the OSA rule is no worse, and in many cases, better than Bessler's rule in terms of ASN. In only one case it was found that the P*-level was not attained, and this occurred for both experiment rules. In general, these two rules gave superior results to those for the pairwise non-design rule, as might be expected. The value of the prior probability g (which was restricted to the range $0.5 \leqslant g < 1$ without loss of generality) appears to have no obvious effect on the relative performance of the rules, which is perhaps surprising, because the OSA rule will randomise between the two experiments as soon as the posterior probability is sufficiently far away from the value 0.5. The values of ν_1 and ν_2 were chosen to yield moderate sample sizes, as this is the situation of interest; in the last case ($\nu_1 = .5$, $\nu_2 = .9$), consideration of performance in terms of efficiency rather than ASN, as discussed above, indicates that the OSA rule is no worse than the other two rules. Finally, it should be noted that the rule of Blot and Meeter (1973) coincides with that of Bessler for the case of two populations.

A study of the problem of 3 binomial populations, in which $\Theta = \{(\nu_2,\nu_1,\nu_1), (\nu_1,\nu_2,\nu_1), (\nu_1,\nu_1,\nu_2)\}$ as in Chapter 3 was not carried out because of the large amount of computer time needed. However, the few results obtained indicated that the OSA rule did not perform well when compared with Bessler's rule. Further investigation indicated that the symmetry of the $\underline{\theta}$ values and the discrete nature of the data led to randomised experimentation under the OSA rule at every stage; i.e. the maximum in (4.37) was attained by more than one experiment at every stage. This is a shortcoming which Bessler's rule does not possess, in general.

Therefore, it may be concluded that the OSA experiment rule is a good rule to use for the problem of identifying one of two single hypotheses concerning the parameters of two binomial populations. This

is indicated by both the optimality property (4.70) and the simulation results in Table 4.2 for problems when a moderate sample size is expected. However, the overshoot of the stopping boundaries and the possibility of randomised experimentation does complicate the comparison of $\underline{\gamma}^0$ with other rules. This is largely due to the discrete nature of the observations themselves; in the next section, consideration of a similar problem involving the parameters of two normal populations, in which the observations are continuous, provides even stronger empirical evidence in favour of $\underline{\gamma}^0$.

The FORTRAN program used to simulate the performance of the OSA experiment rule is given in Appendix A2.

5. IDENTIFYING ONE OF TWO HYPOTHESES CONCERNING THE MEANS AND VARIANCES OF TWO NORMAL POPULATIONS

(1) Set-up And Notation

The problem considered is one in which there are two Normal populations Π_1 and Π_2; it is possible to observe i.i.d. r.v.s $X_1, X_2, \ldots$ from Π_1 with common distribution $N(\mu_X, \sigma_X^2)$, and to observe i.i.d. r.v.s $Y_1, Y_2, \ldots$ from Π_2 with common distribution $N(\mu_Y, \sigma_Y^2)$, where each X_i is independent of each Y_j, $i = 1,2,\ldots$, $j = 1,2,\ldots$, (see (3.12)). The state space $\Theta = \{\underline{\theta}_1, \underline{\theta}_2\}$ is defined by

$$\left.\begin{aligned} \underline{\theta}_1 &= (\mu_2, \sigma_2^2, \mu_1, \sigma_1^2) \\ \underline{\theta}_2 &= (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \end{aligned}\right\}, \quad 0 < \sigma_1^2, \sigma_2^2 < \infty, \quad \mu_1, \mu_2 \in R$$
$$\mu_1 \neq \mu_2, \ \sigma_1^2 \neq \sigma_2^2, \quad \text{all } \underline{\text{known}} \qquad (4.94)$$

where

$$\underline{\theta} = (\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2) \ . \qquad (4.95)$$

Thus the identification involves two components, the mean and the variance

of the populations. As before, the partition $\Theta = \{\Theta_1, \Theta_2\}$ is defined by $\Theta_i = \{\underline{\theta}_i\}$, $i = 1,2$, and the known prior probability that $\underline{\theta}$ equals $\underline{\theta}_1$ is again denoted by g, $0 < g < 1$.

As the population variances σ_X^2 and σ_Y^2 are now components of $\underline{\theta}$, it follows from (3.14) and (3.15) that the sequence of statistics $\{T_n\}$ defined by

$$T_n = \left(n_1, n_2, \sum_{i=1}^{n_1} X_i, \sum_{j=1}^{n_2} Y_j, \sum_{i=1}^{n_1} X_i^2, \sum_{j=1}^{n_2} Y_j^2 \right), \quad n = 1,2,... \tag{4.96}$$

is PARS and POLS for this sequential design problem. (As usual, m and n will be used to denote the number of observations from Π_1 and Π_2, respectively, in this section).

To show that Conditions a) and b) of Theorem 4.4 are satisfied, first note that

$$I(\underline{\theta}_1, \underline{\theta}_2, 1) = \log\left(\frac{\sigma_1}{\sigma_2}\right) - \frac{1}{2} + \frac{\sigma_2^2}{2\sigma_1^2} + \frac{(\mu_2 - \mu_1)^2}{2\sigma_1^2}$$

and

$$I(\underline{\theta}_1, \underline{\theta}_2, 2) = \log\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{1}{2} + \frac{\sigma_1^2}{2\sigma_2^2} + \frac{(\mu_2 - \mu_1)^2}{2\sigma_2^2} \tag{4.97}$$

with $I(\underline{\theta}_2, \underline{\theta}_1, i) = I(\underline{\theta}_1, \underline{\theta}_2, 3 - i)$, $i = 1,2$. (cf. Bessler (1960) part 2, p. 37). Since it is assumed only that σ_1 does not equal σ_2, it follows from the resulting symmetry that Condition a) is true if it can be shown that, say

$$\frac{\sigma_2}{\sigma_1} \exp\left\{\frac{1}{2}\left(\frac{\sigma_1^2}{\sigma_2^2} - 1\right)\right\} \exp\left\{\frac{(\mu_2 - \mu_1)^2}{2\sigma_2^2}\right\} \tag{4.98}$$

is greater than 1. To do this, consider the function $F(x)$ defined on $(0, \infty)$ by

$$F(x) = x^{-1} e^{\frac{1}{2}(x^2 - 1)}. \tag{4.99}$$

It follows that

$$F'(x) = \frac{x^2 - 1}{x^2} e^{\frac{1}{2}(x^2 - 1)}$$

$$\begin{cases} < 0 & \text{if } 0 < x < 1 \\ = 0 & \text{if } x = 1 \\ > 0 & \text{if } x > 1 . \end{cases} \tag{4.100}$$

Therefore, $x = 1$ gives a minimum value of F on $(0, \infty)$, with $F(1) = 1$. As it is assumed that $\sigma_1 \neq \sigma_2$, and clearly $\exp\left\{\frac{(\mu_2 - \mu_1)^2}{2\sigma_2^2}\right\}$ is greater than 1, it follows that the expression (4.98) is always greater than 1, and hence Condition a) is satisfied.

To show that Condition b) holds, note that, e.g.

$$\frac{f_{\underline{\theta}_2}(x|1)}{f_{\underline{\theta}_1}(x|1)} = \frac{\sigma_2}{\sigma_1} \exp\left\{ - \frac{1}{2}\left[\left(\frac{x - \mu_1}{\sigma_1}\right)^2 - \left(\frac{x - \mu_2}{\sigma_2}\right)^2\right]\right\} . \tag{4.101}$$

Therefore, Condition b) follows from the existence of the first four moments of the Normal distribution, and hence all experiment rules for this problem belong to Γ'.

(2) Construction Of The OSA Experiment Rule

It follows from (3.14) that the posterior probability that $\underline{\theta}$ equals $\underline{\theta}_1$, given (m,n) observations $(\underline{X}^m, \underline{Y}^n) = (\underline{x}^m, \underline{y}^n)$, is

$$g(\underline{\theta}_1 | \underline{x}^m, \underline{y}^n) = \left\{1 + \frac{(1 - g) f(\underline{x}^m, \underline{y}^n | \underline{\theta}_2)}{g f(\underline{x}^m, \underline{y}^n | \underline{\theta}_1)}\right\}^{-1}$$

$$= \left\{1 + \left(\frac{1 - g}{g}\right) \sigma_1^{n-m} \sigma_2^{m-n} \exp\left\{ - \frac{1}{2\sigma_1^2\sigma_2^2}\right.\right.$$

$$\times \left[(\sigma_2^2 - \sigma_1^2)\left(\sum_{i=1}^{m} x_i^2 - \sum_{j=1}^{n} y_j^2\right)\right.$$

$$- 2(\sigma_1^2\mu_2 - \sigma_2^2\mu_1)\left(\sum_{j=1}^{n} y_j - \sum_{i=1}^{m} x_i\right)$$

$$\left.\left.\left. + (n - m)(\sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2)\right]\right\}\right\}^{-1} \tag{4.102}$$

which is of the form $(1+a)^{-1}$; similarly, $g(\underline{\theta}_2|\underline{x}^m, \underline{y}^n)$ is of the form $(1+a^{-1})^{-1}$, which will not be written out in full. Thus, referring to (4.38), we are interested in those values of X (say) for which

$$g(\underline{\theta}_1|\underline{x}^m, X, \underline{y}^n) > g(\underline{\theta}_2|\underline{x}^m, X, \underline{y}^n) . \tag{4.103}$$

Direct computation of this inequality yields the following; let α_1 and α_2 $(\alpha_1 \leqslant \alpha_2)$ be the real roots of the quadratic equation

$$(\sigma_2^2-\sigma_1^2)x^2 + 2(\sigma_1^2\mu_2-\sigma_2^2\mu_1)x + 2\sigma_1^2\sigma_2^2\left[\log\left(\frac{g}{1-g}\right) + (n-m-1)\log\left(\frac{\sigma_2}{\sigma_1}\right)\right]$$
$$+ (\sigma_2^2-\sigma_1^2)\left(\sum_{i=1}^{m} x_i^2 - \sum_{j=1}^{n} y_j^2\right) - 2(\sigma_1^2\mu_2-\sigma_2^2\mu_1)\left(\sum_{j=1}^{n} y_j - \sum_{i=1}^{m} x_i\right)$$
$$+ (n-m-1)(\sigma_1^2\mu_2^2-\sigma_2^2\mu_1^2) = 0 . \tag{4.104}$$

Then, (4.103) is true iff

$$\begin{cases} X < \alpha_1 \text{ or } X > \alpha_2 & \text{when } \sigma_2^2 > \sigma_1^2 \\ \alpha_1 < X < \alpha_2 & \text{when } \sigma_2^2 < \sigma_1^2 . \end{cases} \tag{4.105}$$

Similarly, let β_1 and β_2 $(\beta_1 \leqslant \beta_2)$ be the real roots of the quadratic equation

$$(\sigma_2^2-\sigma_1^2)y^2 + 2(\sigma_1^2\mu_2-\sigma_2^2\mu_1)y + 2\sigma_1^2\sigma_2^2\left[\log\left(\frac{1-g}{g}\right) + (m-n-1)\log\left(\frac{\sigma_2}{\sigma_1}\right)\right]$$
$$+ (\sigma_2^2-\sigma_1^2)\left(\sum_{j=1}^{n} y_j^2 - \sum_{i=1}^{m} x_i^2\right) - 2(\sigma_1^2\mu_2-\sigma_2^2\mu_1)\left(\sum_{i=1}^{m} x_i - \sum_{j=1}^{n} y_j\right)$$
$$+ (m-n-1)(\sigma_1^2\mu_2^2-\sigma_2^2\mu_1^2) = 0 . \tag{4.106}$$

Then

$$g(\underline{\theta}_1|\underline{x}^m, \underline{y}^n, Y) > g(\underline{\theta}_2|\underline{x}^m, \underline{y}^n, Y) \quad \text{iff}$$

$$\begin{cases} \beta_1 < Y < \beta_2 & \text{when } \sigma_2^2 > \sigma_1^2 \\ Y < \beta_1 \text{ or } Y > \beta_2 & \text{when } \sigma_2^2 < \sigma_1^2 . \end{cases} \tag{4.107}$$

Therefore, it follows from Theorem 4.6 below that

$$E\left[\max_{i=1,2}\{g(\underline{\theta}_i|\underline{x}^m, X, \underline{y}^n)\}|\underline{x}^m, \underline{y}^n\right]$$

$$= \begin{cases} g(\underline{\theta}_1|\underline{x}^m, \underline{y}^n)\left[\Phi\left(\frac{\alpha_1 - \mu_2}{\sigma_2}\right) + \Phi\left(\frac{\mu_2 - \alpha_2}{\sigma_2}\right)\right] \\ \quad + g(\underline{\theta}_2|\underline{x}^m, \underline{y}^n)\left[\Phi\left(\frac{\alpha_2 - \mu_1}{\sigma_1}\right) - \Phi\left(\frac{\alpha_1 - \mu_1}{\sigma_1}\right)\right] & \text{if } \sigma_2^2 > \sigma_1^2 \\ g(\underline{\theta}_1|\underline{x}^m, \underline{y}^n)\left[\Phi\left(\frac{\alpha_2 - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{\alpha_1 - \mu_2}{\sigma_2}\right)\right] \\ \quad + g(\underline{\theta}_2|\underline{x}^m, \underline{y}^n)\left[\Phi\left(\frac{\alpha_1 - \mu_1}{\sigma_1}\right) + \Phi\left(\frac{\mu_1 - \alpha_2}{\sigma_1}\right)\right] & \text{if } \sigma_2^2 < \sigma_1^2 \end{cases} \tag{4.108}$$

and

$$E\left[\max_{i=1,2}\{g(\underline{\theta}_i|\underline{x}^m, \underline{y}^n, Y)\}|\underline{x}^m, \underline{y}^n\right]$$

$$= \begin{cases} g(\underline{\theta}_1|\underline{x}^m, \underline{y}^n)\left[\Phi\left(\frac{\beta_2 - \mu_1}{\sigma_1}\right) - \Phi\left(\frac{\beta_1 - \mu_1}{\sigma_1}\right)\right] \\ \quad + g(\underline{\theta}_2|\underline{x}^m, \underline{y}^n)\left[\Phi\left(\frac{\beta_1 - \mu_2}{\sigma_2}\right) + \Phi\left(\frac{\mu_2 - \beta_2}{\sigma_2}\right)\right] & \text{if } \sigma_2^2 > \sigma_1^2 \\ g(\underline{\theta}_1|\underline{x}^m, \underline{y}^n)\left[\Phi\left(\frac{\beta_1 - \mu_1}{\sigma_1}\right) + \Phi\left(\frac{\mu_1 - \beta_2}{\sigma_1}\right)\right] \\ \quad + g(\underline{\theta}_2|\underline{x}^m, \underline{y}^n)\left[\Phi\left(\frac{\beta_2 - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{\beta_1 - \mu_2}{\sigma_2}\right)\right] & \text{if } \sigma_2^2 < \sigma_1^2 \end{cases} \tag{4.109}$$

Remark: If either of the quadratic equations (4.104) and (4.106) has no real roots (or one repeated root) then clearly one posterior probability is greater than the other for all possible outcomes of the random variable in question, and hence the expected maximum is equal to that larger posterior probability as follows:

$$\begin{aligned} &\sigma_2^2 > \sigma_1^2 : g(\underline{\theta}_1|\underline{x}^m, \underline{y}^n) \text{ for X}, \; g(\underline{\theta}_2|\underline{x}^m, \underline{y}^n) \text{ for Y} \\ &\sigma_2^2 < \sigma_1^2 : g(\underline{\theta}_2|\underline{x}^m, \underline{y}^n) \text{ for X}, \; g(\underline{\theta}_1|\underline{x}^m, \underline{y}^n) \text{ for Y}. \end{aligned} \tag{4.110}$$

Therefore, the two expectations (4.108) and (4.109) can never be the same in the sense of (4.110) (except when $g(\underline{\theta}_1|\cdot) = g(\underline{\theta}_2|\cdot) = \cdot 5$), and hence

randomised experimentation under $\underline{\gamma}^0$ will almost certainly never occur.

Theorem 4.6 In the notation of Chapter 2, define

$$X_{i,n} = \{x \in X : g(\underline{\theta}_i|\underline{x}^n, x, \underline{e}^{n+1}) = \max_{1\leqslant j\leqslant k}\{g(\underline{\theta}_j|\underline{x}^n, x, \underline{e}^{n+1})\}\} \quad (4.111)$$

(assuming ties occur with probability 0). Then

$$E[\max_{1\leqslant j\leqslant k}\{g(\underline{\theta}_j|\underline{x}^n, X, \underline{e}^{n+1})\}|\underline{x}^n, \underline{e}^{n+1}]$$
$$= \sum_{j=1}^{k} g(\underline{\theta}_j|\underline{x}^n, \underline{e}^n) \cdot \int_{X_{j,n}} f_{\underline{\theta}_j}(x|e_{n+1})dx \quad (4.112)$$

Proof: By definition

$$E[\max_{1\leqslant j\leqslant k}\{g(\underline{\theta}_j|\underline{x}^n, X, \underline{e}^{n+1})\}|\underline{x}^n, \underline{e}^{n+1}]$$
$$= \sum_{j=1}^{k} \int_{X_{j,n}} g(\underline{\theta}_j|\underline{x}^n, x_{n+1}, \underline{e}^{n+1}) f_G(x_{n+1}|\underline{x}^n, \underline{e}^{n+1})dx_{n+1}$$

where $f_G(x_{n+1}|\underline{x}^n, \underline{e}^{n+1})$ was defined in (2.45),

$$= \sum_{j=1}^{k} \int_{X_{j,n}} \frac{f_{\underline{\theta},\underline{\gamma}^n}(\underline{x}^{n+1},\underline{e}^{n+1})g(\underline{\theta})}{f_{G,\underline{\gamma}^n}(\underline{x}^{n+1},\underline{e}^{n+1})} \cdot \left[\int_\Theta f_{\underline{\theta}}(x_{n+1}|\underline{x}^n,\underline{e}^{n+1})g(\underline{\theta}|\underline{x}^n,\underline{e}^n)d\nu(\theta)\right] dx_{n+1}$$

from (2.39) and (2.45),

$$= \sum_{j=1}^{k} \int_{X_{j,n}} \left[\frac{f_{\underline{\theta}}(x_{n+1}|\underline{x}^n,\underline{e}^{n+1}).\gamma_n(e_{n+1}|\underline{x}^n,\underline{e}^n) f_{\underline{\theta},\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n)g(\underline{\theta})}{f_{G,\underline{\gamma}^n}(\underline{x}^{n+1},\underline{e}^{n+1})}\right]$$
$$\times \left[\frac{f_{G,\underline{\gamma}^n}(\underline{x}^{n+1},\underline{e}^{n+1})}{\gamma_n(e_{n+1}|\underline{x}^n,\underline{e}^n).f_{G,\underline{\gamma}^{n-1}}(\underline{x}^n,\underline{e}^n)}\right] dx_{n+1}$$

from (2.39) and (2.45) again,

$$= \sum_{j=1}^{k} g(\underline{\theta}_j|\underline{x}^n, \underline{e}^n) \cdot \int_{X_{j,n}} f_{\underline{\theta}}(x_{n+1}|e_{n+1})dx_{n+1} \quad (4.113)$$

since $f_{\underline{\theta}}(x_{n+1}|\underline{x}^n, \underline{e}^{n+1}) \equiv f_{\underline{\theta}}(x_{n+1}|e_{n+1})$ here. □

(3) The Performance Of $\underline{\gamma}^0$

Table 4.3 lists the performance of the OSA experiment rule, Bessler's rule, and the pairwise non-design rule for selected values of g, σ_1 and σ_2 under computer simulation. It may easily be shown that the

TABLE 4.3. Performance of OSA (first entry), Bessler's (second entry) and pairwise (third entry) experiment rules for certain values of g, σ_1, σ_2.

$P^* = 0{\cdot}95$. $\mu_2 - \mu_1 = 0{\cdot}5$. 400 simulations.

g	σ_1	σ_2	P(CS)	ASN	Standard Error (ASN)	Theoretical ASN
.5	1	2	.970	7.11	0.20	4.14
			.995	7.42	0.29	
			.993	8.22	0.30	
.5	2	3	.963	16.8	0.47	13.5
			.990	20.1	0.84	
			.980	21.2	0.79	
.5	2	4	.970	7.62	0.24	4.55
			.990	7.71	0.32	
			.988	8.88	0.30	
.7	1	2	.975	6.30	0.18	4.10
			.985	7.26	0.30	
			.980	7.51	0.29	
.7	2	3	.963	15.6	0.50	13.5
			.988	17.4	0.75	
			.978	19.2	0.72	
.7	2	4	.965	6.98	0.22	4.55
			.985	7.24	0.30	
			.983	7.79	0.28	

TABLE 4.3. continued

g	σ_1	σ_2	P(CS)	ASN	Standard Error (ASN)	Theoretical ASN
.5	3	4	.960	30.6	1.22	27.6
			.980	33.9	1.37	
			.970	35.5	1.29	
.5	3	5	.970	12.7	0.35	9.02
			.983	13.8	0.53	
			.983	14.7	0.53	
.5	4	5	.951	49.0	2.02	46.5
			.939	55.9	2.22	
			.970	55.3	2.04	
.7	3	4	.958	29.0	1.04	27.6
			.970	33.3	1.39	
			.965	34.2	1.34	
.7	3	5	.965	11.1	0.36	9.02
			.983	12.5	0.54	
			.988	13.6	0.52	
.7	4	5	.945	44.3	2.12	46.5
			.933	49.4	1.70	
			.970	53.6	2.03	

likelihood ratio of the two sampling densities, and hence the posterior densities themselves, depend on the population means only through the difference $\mu_2 - \mu_1$; this difference was set equal to 0·5, and then values of σ_1 and σ_2 were chosen to ensure moderate sample sizes.

From these results, it is seen that the OSA rule is superior to both Bessler's rule and the pairwise rule in every case considered; in

only one case was the significance level not attained, and this again occurred for both design rules. Much of this superiority seems to be due to the consistent overshoot of the stopping boundaries by the other two rules in relation to $\underline{\gamma}^0$, which is somewhat surprising in the case of Bessler's rule. As before, the prior probability g appears to have little effect on relative performance.

Once again, it is possible to obtain an analytical expression for the expected sample size under pairwise sampling ignoring overshoot again. Proceeding as before, the expression

$$E[N] = \frac{2\sigma_1^2\sigma_1^2\left[(2P^* - 1)\log\left(\frac{P^*}{1-P^*}\right) + (2g-1)\log\left(\frac{1-g}{g}\right)\right]}{(\sigma_1^2+\sigma_2^2).(\mu_2-\mu_1)^2 + (\sigma_1^2-\sigma_2^2)^2} \qquad (4.114)$$

is listed in Table 4.3 with $\mu_2 - \mu_1 = 0.5$ and $P^* = 0.95$. It is seen that, in every case where the P*-level is attained, the observed ASN is greater than the theoretical ASN. Once again, this discrepancy may be explained by overshoot of the boundaries and error in simulation. Such discrepancies have been noted in similar studies; see Robbins and Siegmund (1974), Tables 1 and 2, for example.

Therefore, it may be concluded that although all three design rules appear to exceed the P*-level consistently in simulation, the OSA experiment rule has a smaller exceedance than the other rules in almost every case considered, and hence gives rise to a smaller ASN value. Moreover, these conclusions appear to hold true regardless of the value of the prior probability g.

The FORTRAN program used to simulate the performance of the OSA experiment rule is given in Appendix A2.

6. CONCLUDING REMARKS

The approach taken in this chapter has been practical rather than theoretical; and deliberately so, because the theory of Bayes sequential

design procedures is of limited use to the practitioner, as was seen in Chapter 3. In a sense, the comments made in Chapter 1 concerning asymptotic optimality hold for Bayes sequential design optimality as well; what the experimenter requires is a procedure which is easily applicable in a wide variety of situations and is relatively efficient in terms of risk or sample size. It is well-known that most workers in agriculture and the social sciences etc. tend to opt for the former, often at the expense of the latter. For many, collecting more data is easier than "working out the mathematics."

Nevertheless, any procedure which is proposed as suitable in both respects (by a theoretical statistician, anyway) should be based on some theoretical basis of "goodness", such as asymptotic optimality, if for no other reason than to justify its use with reasons more mathematical than "it seems to work well." No-one expects a rule to be optimal when its structure is quite different from that of the optimal procedure, but some sort of rational justification for its use should be provided. Such is the case, it is claimed, with the OSA experiment rule; for, although it cannot be optimal except in trivial cases, its structure is similar to, and based on, that of the optimal rule; in the two-binomial identification problem, it is shown to possess an optimality property which seems intuitively to be related to efficiency in terms of expected sample size; and finally, the computer results indicate that it indeed "seems to work well."

However, from a theoretical standpoint, results such as these can only serve to indicate the need for future research and the possible directions it may take. As usual, more questions arise during the course of research than are ever answered:

(1) With respect to the modified goal, little if anything is known about the optimal experiment rule (optimal in the sense of (4.20), that is),

or even whether an optimal experiment rule exists or not. Under what, if any, situations is the OSA experiment rule $\underline{\gamma}^0$ optimal?

(2) Notwithstanding remarks made above and in Chapter 1, it would be of interest to know whether $\underline{\gamma}^0$ is asymptotically optimal or not. Sufficient conditions for a rule to be asymptotically optimal are given by Bessler, but unfortunately these are stated in terms of information numbers and payoff matrices, and consequently it is not clear how to apply them in this case. In fact, even with a rule such as Blot and Meeter's which is based on the concepts, it is found that most of the paper is devoted to showing that their rule satisfies these conditions.

(3) It would be of benefit if it could be shown that the optimality property (4.70) does imply a minimisation of the expected sample size in some sense; also, the computer results in the Normal case indicate that a similar property might be proved in that case.

(4) Given time, and especially computer time, more extensive computer studies in a larger variety of situations, especially for $k \geqslant 3$ populations, could provide insight into the behaviour of competing rules and that of the optimal rule itself (if it exists).

Finally, it may be asked why the performance of the OSA rule has not been considered in two obvious cases; a) the analogue of the Normal example above, with (4.94) modified by $\sigma_1^2 = \sigma_2^2$ (i.e. equal variances, different means), and b) the analogue of the Normal ranking sample in Chapter 3, with two independent Normal prior distributions and 0 - 1 loss. In the first case, it has been shown by Robbins and Siegmund (1974) that the pairwise non-design rule minimise the expected sample size, and it may be shown that the OSA rule will randomise at every stage here; that is, the expectations in (4.38) are always equal. In the second case, it was found that under 0 - 1 loss, even the first stage of backward induction was intractible; consequently, the one-step-ahead expectations cannot be evaluated.

ACKNOWLEDGEMENTS

I would like to express my thanks to Professor J.J. Deely for his supervision and advice during the preparation of this thesis. I am also grateful to Kelly Mara and Dr. Murray Smith for their many helpful comments, and to Ann Tindall for such an excellent job of typing the manuscript.

REFERENCES

ALBERT, A.E. (1961). The sequential design of experiments for infinitely many states of nature. Ann. Math. Statist. 32 774-799.

ARROW, K.J., BLACKWELL, D., and GIRSHICK, M.A. (1949). Bayes and minimax solutions of sequential decision problems. Econometrica 17 213-244.

BECHHOFER, R.E., KIEFER, J., and SOBEL, M. (1968). Sequential identification and ranking procedures. Chicago, University of Chicago Press. 420p.

BERRY, D.A. (1972). A Bernoulli two-armed bandit. Ann. Math. Statist. 43 871-897.

BERRY, D.A. (1978). Modified two-armed bandit strategies for certain clinical trials. J. Amer. Statist. Assoc. 73 339-345.

BESSLER, S.A. (1960. Theory and applications of the sequential design of experiments, k-actions and infinitely many experiments. Technical Report No.s 55 and 56, Applied Mathematics And Statistics Laboratories, Stanford University, Stanford, California.

BLOT, W.J., and MEETER, D.A. (1973). Sequential experimental design procedures. J. Amer. Statist. Assoc. 68 586-593.

CHERNOFF, H. (1959). Sequential design of experiments. Ann. Math. Statist. 30 755-770.

CHERNOFF, H. (1972). Sequential analysis and optimal design. Regional Conference Series In Applied Mathematics No. 8, Society For Industrial And Applied Mathematics, Philadelphia, Pennsylvania.

CHERNOFF, H. (1975). Approaches in sequential design of experiments. In A Survey of statistical design and linear models, ed. J.N. Srivastava; Amsterdam, North-Holland Publishing Company, p. 67-90.

CHOW, Y.S., ROBBINS, H., and SIEGMUND, D. (1971). Great Expectations: the theory of optimal stopping. Boston, Houghton Mifflin. 138p.

DeGROOT, M.H. (1970). Optimal statistical decisions. New York, McGraw-Hill. 489p.

FELDMAN, D. (1962). Contributions to the "two-armed bandit" problem. Ann. Math. Statist. 33 847-856.

FELLER, W. (1966). An introduction to probability theory and its applications, Volume 2. New York, John Wiley & Sons. 626p.

FERGUSON, T.S. (1967). Mathematical statistics: a decision theoretic approach. New York, Academic. 396p.

FISZ, M. (1963). Probability theory and mathematical statistics. 3d. ed. New York, John Wiley & Sons. 677p.

GRAY, K.B., Jr. (1968). Sequential selection of experiments. Ann. Math. Statist. 39 1953-1977.

HAGGSTROM, G.W. (1966). Optimal stopping and experimental design. Ann. Math. Statist. 37 7-29.

HOEL, D.G., SOBEL, M., and WEISS, G.H. (1975). A survey of adaptive sampling for clinical trials. In Perspectives in biometrics, Volume I, ed. R.M. Elashoff. New York, Academic, p. 29-61.

KIEFER, J., and SACKS, J. (1963). Asymptotically optimum sequential inference and design. Ann. Math. Statist. 34 705-750.

KINGMAN, J.F.C., and TAYLOR, S.J. (1966). Introduction to measure and probability. London, Cambridge University Press. 401p.

LINDLEY, D.V. (1965). Introduction to probability and statistics from a Bayesian viewpoint. Part 2: Inference. London, Cambridge University Press. 292p.

ROBBINS, H. (1952). Some aspects of the sequential design of experiments. Bull. Amer. Math. Soc. 55 527-535.

ROBBINS, H. (1974). A sequential test for two binomial populations. Proc. Nat. Acad. Sci. USA 71 4435-4436.

ROBBINS, H., and SIEGMUND, D. (1974). Sequential tests involving two populations. J. Amer. Statist. Soc. 69 132-139.

SCHWARZ, G. (1962). Asymptotic shapes of Bayes sequential testing regions. Ann. Math. Statist. 33 224-236.

WALD, A. (1947). Sequential analysis. New York, John Wiley & Sons. 212p.

WALD, A. (1950). Statistical decision functions. New York, John Wiley & Sons. 179p.

APPENDIX A1

EXTENSION OF HAGGSTROM'S RESULTS TO THE CASE WHERE THE EXPERIMENT SPACE IS INFINITE

In this section, the notation of Haggstrom (1966) is used to extend the two principal results of that paper to the case where the experiment space may be infinite.

Let $M = \{1,2,\ldots, m\}$ be the (finite) set of experiments. Let A^m be the set of all finite sequences of elements of M, including the sequence of "no components", which is denoted as ϕ.

Define the partially-ordered set $\{A^m, \leqslant\}$ by

$$a = (a_1,\ldots, a_j) \leqslant b = (b_1,\ldots, b_i)$$
$$\text{iff} \quad j \leqslant i \quad \text{and} \quad a_k = b_k \quad \text{for all } k \leqslant j. \tag{A1.1}$$

i.e. $a \leqslant b$ iff a is an "initial segment" of b.

Let A^{m*} be the set of infinite sequences of elements of M. Then the partially ordered set $\{A^m \cup A^{m*}, \leqslant\}$ is defined by:

$$\text{for} \quad a,b \in A^m ; \qquad a \leqslant b \quad \text{as in (A1.1)}$$
$$\text{for} \quad a \in A^m,\ b \in A^{m*} ; \qquad a = (a_1,\ldots, a_j) \leqslant b$$
$$\text{iff} \quad a_k = b_k \quad \text{for all } k \leqslant j. \tag{A1.2}$$

Let $\{Z_a, F_a, a \in A^m\}$ be a non-negative (a.e.), integrable stochastic process on a probability space $\{\Omega, F, P\}$, i.e.

(1) The Fa's are σ-fields of subsets of Ω, and $F_a \subseteq F_b \subseteq F$ for all $a \leqslant b$, $a,b \in A^m$.

(2) Each Z_a is F_a-measurable, integrable, and non-negative (a.e.).

Using the notation $ak = (a_1,\ldots, a_j, k)$, $a \in A^m$, $k \in M$, and $ab = (a_1,\ldots, a_j, b_1,\ldots)$, $a \in A^m$, $b \in A^{m*}$, the set T of control variables

t is defined, where a control variable (c.v.) t is a mapping from Ω to $A^m \cup A^{m*}$ satisfying

$$\begin{aligned} &(1) \quad t \in A^m \quad \text{a.e.} \\ &(2) \quad \{t = a\} \quad \text{and} \quad \{t \geqslant ak\} \in \mathcal{F}_a \quad \text{for all } a \in A^m. \end{aligned} \tag{A1.3}$$

For a c.v.t, define Z_t to be a r.v. satisfying

$$Z_t(\omega) = \begin{cases} Z_a(\omega) & \text{if} \quad t(\omega) = a \in A^m \\ \infty & \text{if} \quad t(\omega) = b \in A^{m*} \end{cases} . \tag{A1.4}$$

The optimal c.v. t*, if it exists, satisfies

$$E[Z_{t*}] \leqslant E[Z_t] \quad \text{for all } t \in T . \tag{A1.4}$$

Remarks

(1) The sequences $a = (a_1, \ldots, a_j) \in A^m$ and $b \in A^{m*}$ correspond to the notation $\underline{e}^j = (e_1, \ldots, e_j)$ and $\underline{e} = (e_1, e_2, \ldots)$, respectively. Similarly, the σ-fields $\mathcal{F}_a$ and $\mathcal{F}_b$ correspond to $\mathcal{F}_n$ and $\mathcal{F}$, respectively (see (2.6) and (2.9)).

(2) Z_a corresponds to the expected posterior loss given $(\underline{X}^j, \underline{E}^j) = (\underline{x}^j, \underline{e}^j)$, namely $U_j(g_j)$, defined in (2.44).

(3) The control variable t corresponds to the stopping time N associated with the stopping rule $\underline{\phi}$, whose distribution depends on the experiment rule $\underline{\gamma}$ (see (2.67) and (2.68)). Thus t incorporates both the stopping policy and the experiment policy, without allowing the possibility of a randomised experiment choice. As was seen in Chapter 2, however, the optimal experiment rule may be taken to be nonrandomised.

Infinite Experiment Space

All the preceding definitions etc. still hold if the sets A^m and A^{m*} are replaced throughout by A^∞ and $A^{\infty*}$ respectively, where

$$A^{\infty} \equiv \text{set of all finite sequences of elements of } M^*$$

$$A^{\infty *} \equiv \text{set of all infinite sequences of elements of } M^* \qquad \text{(A1.5)}$$

where $M^* = \{1,2,\ldots\}$.

It is not the purpose of this thesis to reproduce Haggstrom's paper; in what follows, the relevant results in that paper are shown to hold when $\mathfrak{E}$ is countably infinite as described above.

Lemma A1.1 Given a class of r.v.s $\{x_i, i \in I\}$, where the index set I is a subset of the real line; if $I = \bigcup_{j=1}^{\infty} I_j$, then

$$\operatorname*{ess\,inf}_{i \in I} x_i = \inf_j \operatorname*{ess\,inf}_{i \in I_j} x_i \qquad \text{(A1.6)}$$

Proof: a) $I_j \subset I$ for all j

$$\Rightarrow \quad \operatorname*{ess\,inf}_{i \in I} x_i \leqslant \operatorname*{ess\,inf}_{i \in I_j} x_i \quad \text{for all } j$$

$$\Rightarrow \quad \operatorname*{ess\,inf}_{i \in I} x_i \leqslant \inf_j \operatorname*{ess\,inf}_{i \in I_j} x_i$$

b) By definition of essential infimum,

$$\operatorname*{ess\,inf}_{i \in I} x_i = \inf_{t \in T} x_t \quad \text{for some countable subset T of I}$$

and for each j,

$$\operatorname*{ess\,inf}_{i \in I_j} x_i = \inf_{t \in T_j} x_t \quad \text{for some countable subset } T_j \text{ of } I_j .$$

Without loss of generality, each T_j may be chosen such that $T \cap I_j \subseteq T_j$. Therefore, for each $j = 1,2,\ldots,$

$$\inf_j \inf_{t_j \in T_j} x_{t_j} \leqslant \inf_{t_j \in T_j} x_{t_j}$$

and

$$\inf_{t_j \in T_j} x_{t_j} \leqslant x_t \quad \text{for all } t \in T \cap I_j .$$

Combining the last two inequalities, it follows that

$$\inf_j \inf_{t_j \in T_j} x_{t_j} \leqslant x_t \quad \text{for all } t \in T \cap I_j, \quad \text{for all } j = 1,2,\ldots$$

$$\Rightarrow \quad \inf_j \inf_{t_j \in T_j} x_{t_j} \leqslant \inf_{t \in T} x_t$$

and the result follows. □

The result corresponding to Lemma 4.1 (c) of Haggstrom follows from Lemma A1.1, namely

Lemma A1.2 $$X_a = \min\{Z_a, \inf_{j \in M^*} E[X_{aj} | \mathcal{F}_a]\} . \tag{A1.7}$$

It is now possible to prove Theorem 2.4 concerning the optimality of the experiment rule γ^*; this follows from Theorem A1.1 below (see comments on pp. 21-22 of Haggstrom (1966)).

Theorem A1.1. Theorem 4.1 of Haggstrom (1966) holds if A^m is replaced by A^∞.

Proof: This parallels the proof of Haggstrom's Theorem 4.1. Let A_n^m (respectively, A_n^∞) denote the set of elements of A^m (respectively, A^∞) having exactly n components. Further, let $t^*(\omega) < A_n^\infty$ mean that $t^*(\omega) < a$ for some $a \in A_n^\infty$ etc. For each $n = 0,1,2,\ldots$, define

$$\left.\begin{aligned} f_{n,m} &= \sum_{a \in A_n^m} I_{\{t^* \geqslant a\}} X_a \\ f_n &= \sum_{a \in A_n^\infty} I_{\{t^* \geqslant a\}} X_a \end{aligned}\right\} \tag{A1.8}$$

where I_F is the indicator function of F, $F \in \mathcal{F}$.

Then

$$f_n - f_{n,m} = \sum_{\substack{\{a \in A_n^\infty : \text{each } a_i \\ \in \{m+1, m+2, \ldots\}\}}} I_{\{t^* \geqslant a\}} X_a \to 0 \quad \text{as} \quad m \to \infty$$

i.e. for each $n = 0,1,\ldots$, $\{f_{n,m}\}_{m=1}^{\infty}$ is an increasing sequence of $\mathcal{F}_a$-measurable and integrable functions, and $f_{n,m} \to f_n$ as $m \to \infty$. Thus, f_n is $\mathcal{F}_a$-measurable and integrable by the monotone convergence theorem. Similarly, it may be shown that $I_{\{t^* < A_n^{\infty}\}} Z_{t^*}$ is $\mathcal{F}_a$-measurable and integrable. Therefore, the conditional expectation

$$X_{\phi} = E\left[I_{\{t^* < A_n^{\infty}\}} Z_{t^*} + \sum_{a \in A_n^{\infty}} I_{\{t^* \geq a\}} X_a \middle| \mathcal{F}_{\phi}\right] \tag{A1.9}$$

is well-defined for each $n = 0,1,\ldots$. The proof of (A1.9) by induction on n follows similarly; i.e. for each $n = 0,1,2,\ldots$, define

$$\left.\begin{aligned} g_{n,m} &= \sum_{a \in A_n^m} \sum_{j \in M} I_{\{t^* \geq aj\}} E[X_{aj} | \mathcal{F}_a] \\ g_n &= \sum_{a \in A_n^{\infty}} \sum_{j \in M^*} I_{\{t^* \geq aj\}} E[X_{aj} | \mathcal{F}_a] \end{aligned}\right\} \tag{A1.10}$$

Then $g_{n,m} \uparrow g_n$, g_n is $\mathcal{F}_a$-measurable and integrable, and hence $E[g_{n,m} | \mathcal{F}_{\phi}] \uparrow E[g_n | \mathcal{F}_{\phi}]$ by the monotone convergence theorem. □

Finally, the existence of the optimal stopping rule $\underline{\phi}^*$ and the optimal experiment rule $\underline{\gamma}^*$ follows from Theorem A1.2 below.

Theorem A1.2. Theorem 4.3 of Haggstrom (1966) holds if A^m is replaced by A^{∞}.

Proof: This follows in an analagous fashion to the proof of Theorem A1.1. Referring to the proof of Theorem 4.3 of Haggstrom, $\{\mathcal{F}_n,\ n \geq 0\}$ is still an increasing sequence of σ-fields if A_n^m is replaced by A_n^{∞}. Denoting Haggstrom's V_n as V_n^m, and now denoting

$$V_n = I_{\{t^* < A_n^{\infty}\}} X_{t^*} + \sum_{a \in A_n^{\infty}} I_{\{t^* \geq a\}} X_a$$

it follows that

$$V_n - V_n^m \leq \sum_{\substack{\{a \in A_n^{\infty} : \text{each } a_i \\ \in \{m+1, m+2, \ldots\}\}}} \left[I_{\{t^* < a\}} X_{t^*} + I_{\{t^* \geq a\}} X_a\right]$$

which decreases monotonically to 0, for any value of $n = 0,1,\ldots$. Thus V_n is, for each n, F_n-measurable and integrable, and

$$\begin{aligned} E[V_{n+1}|F_n] &= E[\lim_{m\to\infty} V^m_{n+1}|F_n] \\ &= \lim_{m\to\infty} E[V^m_{n+1}|F_n] \quad \text{by the monotone convergence theorem} \\ &= \lim_{m\to\infty} V^m_n \quad \text{since } \{V^m_n, F_n, n=0,1,\ldots\} \text{ is a martingale} \\ &\qquad\qquad \text{for any } m \\ &= V_n \end{aligned}$$

i.e. $\{V_n, F_n, n = 0,1,\ldots\}$ is a martingale. The proof then follows as in Haggstrom's Theorem 4.3. □

APPENDIX A2

COMPUTER PROGRAMS

B A C K I N D

```
C     THIS PROGRAM COMPUTES THE BAYES SEQUENTIAL DESIGN PROCEDURE AND
C
C     THE CORRESPONDING BAYES RISK FOR THE SIMPLE IDENTIFICATION PROBLEM
C
C     CONCERNING THE PARAMETERS OF TWO BINOMIAL POPULATIONS UNDER 0-1
C
C     LOSS.
C
C     NO = NUMBER OF OBSERVATIONS OR TRUNCATION VALUE
C
C     NP = NUMBER OF POPULATIONS (TWO IN THIS CASE)
C
C     RF = WEIGHT OF LOSS FUNCTION AGAINST UNIT COST
C
C     C = PRIOR PROBABILITY
C
C     EM1,EM2 = IDENTIFIED PARAMETER VALUES
C

      DIMENSION DG(2),FG(2),ER(2),F(2),R(2),K(4),M(4),N(4),RC(13,13,13,1
     13)
      NO=4
      NP=2
      NV=NO/NP
      RF=50.
      C=.7
      EM1=.2
      EM2=.7
      DEM1=1.-EM1
      DEM2=1.-EM2
      WRITE(6,500)

      WRITE(6,550)
      WRITE(6,600)C,EM1,EM2,NO,RF
      DO 15 IC=1,NO+1
      IP=NO+2-IO
      IL=IP-1
      CALL COMB(IL,NP,KJ)
      WRITE(6,800)IL,KJ
      WRITE(6,900)
      WRITE(6,950)
      DO 15 I1=1,IP
      M(1)=I1-1
      DO 15 I2=1,IP+1-I1
      M(2)=I2-1
      DO 15 I3=1,IP+2-I1-I2
      M(3)=I3-1
      M(4)=IL-M(1)-M(2)-M(3)
      DO 12 J=1,4
      K(J)=M(J)+1
   12 CONTINUE
      F(1)=C*DEM2**M(1)*EM2**M(2)*DEM1**M(3)*EM1**M(4)
      F(2)=(1.-C)*DEM1**M(1)*EM1**M(2)*DEM2**M(3)*EM2**M(4)
      DEN=F(1)+F(2)
      DO 13 IS=1,2
      JS=3-IS
   13 R(IS)=(RF*F(JS))/DEN
      AIL=IL
      RO=AMIN1(R(1),R(2))+AIL
      DO 9 LI=1,2
      IF(R(LI).LE.(RO-AIL))IND=LI
    9 CONTINUE
      IF(IL.EQ.NO)INS=1
      IF(IL.EQ.NO)WRITE(6,100)(M(L),L=1,4),RO,INS,IND,DEN
      IF(IL.EQ.NO)RC(K(1),K(2),K(3),K(4))=RO
      IF(IL.EQ.NO)GO TO 15
      DO 14 IT=1,2
      JT=3-IT
      FG(IT)=(F(IT)*DEM2+F(JT)*DEM1)/DEN
   14 DG(IT)=1.-FG(IT)
C
```

```
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
    START OF SEGMENT 002
FORMAT SEGMENT IS 00CE LONG
C   002:0000:0
C   002:0000:0
C   002:0000:0
C   002:0000:5
C   002:0001:4
C   002:0003:1
C   002:0004:0
C   002:0006:3
C   002:0008:3
C   002:000A:3
C   002:000B:4
C   002:000C:5
       FIB IS 0006 LONG
C   002:0011:2
C   002:0015:2
C   002:0022:2
C   002:0026:3
C   002:0028:3
C   002:0029:3
C   002:002B:4
C   002:0034:2
C   002:0038:2
C   002:003C:2
C   002:003D:0
C   002:003E:4
C   002:0043:0
C   002:0044:4
C   002:0049:4
C   002:004B:3
C   002:004F:1
C   002:0050:0
C   002:0052:5
C   002:0055:0
C   002:005E:3
C   002:0068:2
C   002:006A:0
C   002:006B:0
C   002:006C:3
C   002:0072:0
C   002:0072:5
C   002:0076:3
C   002:0078:0
C   002:007B:2
C   002:007D:3
C   002:007F:3
C   002:0093:2
C   002:009A:4
C   002:009C:0
C   002:009D:0
C   002:009E:3
C   002:00A3:3
C   002:00A8:2
C   002:00A8:2
```

```
      DO 16 IG=1,4
      N(IG)=K(IG)+1
   16 CONTINUE
      ER(1)=FG(1)*RC(N(1),K(2),K(3),K(4))+DG(1)*RC(K(1),N(2),K(3),K(4))
      ER(2)=FG(1)*RC(K(1),K(2),N(3),K(4))+DG(1)*RC(K(1),K(2),K(3),N(4))
      RM=AMIN1(ER(1),ER(2))
      DO 8 LZ=1,2
      IF(ER(LZ).LE.RM)INE=LZ
    8 CONTINUE
      IF(ABS(ER(1)-ER(2)).LE.1.E-4)INE=12
      RC(K(1),K(2),K(3),K(4))=AMIN1(RO,RM)
      IF(RO.GT.RM)INS=0
      IF(RO.LE.RM)INS=1
      IF(INS.EQ.0)WRITE(6,200)(M(L),L=1,4),RO,(ER(I),I=1,2),RC(K(1),K(2)
     1,K(3),K(4)),INS,INE,DEN
      IF(INS.EQ.1)WRITE(6,300)(M(L),L=1,4),RO,(ER(I),I=1,2),RC(K(1),K(2)
     1,K(3),K(4)),INS,IND,DEN
   15 CONTINUE
      WRITE(6,400)RC(1,1,1,1)
  400 FORMAT(////41X,'**********************',/41X,'*BAYES RISK =',F8.4,
     1'*',/41X,'**********************')
  500 FORMAT(////////////////35X,'CONSTRUCTION OF THE BAYES SEQUENTIAL DE
     1SIGN RULE TO',//35X,'IDENTIFY THE PARAMETERS OF TWO BINOMIAL POPUL
     2ATIONS')
  550 FORMAT(//36X,'H1 : THETA = (NU2,NU1)  VS  H2 : THETA = (NU1,NU2)')
  600 FORMAT(///10X,'G=',F5.2,5X,'NU1 =',F5.2,3X,'NU2 =',F5.2,8X,'TRUNCA
     1TED AT',I2,1X,'OBSERVATIONS',8X,'LOSS FACTOR =',F5.1)
  800 FORMAT(/////////20X,'NO. OF OBSERVATIONS =',I2,10X,'NO. OF STATES
     1=',I3,/19X,'**************************',8X,'**********************')
  900 FORMAT(///15X,'*',2X,'STATE',1X,'*',4X,'RISK',4X,'*',1X,'E(RISK(X)
     1)',1X,'*',1X,'E(RISK(Y))',1X,'*',1X,'MIN. RISK',2X,'*STOP',1X,'*',
     21X,'EXP',1X,'*',1X,'DECN',1X,'*',2X,'FG(STATE)',1X,'*')
  950 FORMAT(15X,'**************************************************
     1**********************************')
  200 FORMAT(/15X,'*',2X,4I1,2X,'*',2X,F8.4,2X,'*',2X,F8.4,2X,'*',2X,F8.
     14,2X,'*',2X,F8.4,2X,'*',1X,I3,1X,'*',1X,I3,1X,'*',6X,'*',2X,F8.4,2
     22X,'*')
  300 FORMAT(/15X,'*',2X,4I1,2X,'*',2X,F8.4,2X,'*',2X,F8.4,2X,'*',2X,F8.
     14,2X,'*',2X,F8.4,2X,'*',1X,I3,1X,'*',5X,'*',1X,I3,2X,'*',2X,F8.4,2
     22X,'*')
  100 FORMAT(/15X,'*',2X,4I1,2X,'*',2X,F8.4,2X,'*',12X,'*',12X,'*',12X,'
     1*',1X,I3,1X,'*',5X,'*',1X,I3,2X,'*',2X,F8.4,2X,'*')
  723 STOP
      END
```

```
C  002:00A8:2
C  002:00A8:2
C  002:00A8:2
C  002:00A8:2
C  002:00A8:2
C  002:00A8:2
C  002:00A8:2
C  002:00A8:2
C  002:00A8:2
C  002:00A9:0
C  002:00AB:5
C  002:00AE:0
C  002:00BB:0
C  002:00C8:0
C  002:00C8:1
C  002:00CC:0
C  002:00CE:5
C  002:00D1:0
C  002:00D5:3
C  002:00DD:3
C  002:00DF:3
C  002:00E1:3
C  002:00F5:5
C  002:0102:2
C  002:0116:5
C  002:0123:2
C  002:0127:0
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012D:2
C  002:012E:1
SEGMENT 002 IS 0142 LONG
```

```
      SUBROUTINE COMB(I,N,K)
C
C     THIS SUBROUTINE EVALUATES THE COMBINATORIAL K = (I+2N-1 CHOOSE
C     2N-1) DIRECTLY.  IT IS USED TO COMPUTE THE NUMBER OF POSSIBLE
C     VALUES OF THE SUFFICIENT STATISTIC TI AFTER I STAGES OF
C     EXPERIMENTATION BASED ON N DIFFERENT EXPERIMENTS
C
      M1=I+2*N-1
      M2=2*N-1
      M3=I
      N1=1
      N2=1
      N3=1
      DO 30 M=1,M1
   30 N1=N1*M
      DO 31 M=1,M2
   31 N2=N2*M
      IF(M3.EQ.0)GO TO 33
      DO 32 M=1,M3
   32 N3=N3*M
      GO TO 34
   33 N3=1
   34 K=N1/(N2*N3)
      RETURN
      END
```

```
 START OF SEGMENT 006
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0000:0
C   006:0002:2
C   006:0004:1
C   006:0005:0
C   006:0005:4
C   006:0006:2
C   006:0007:0
C   006:0008:0
C   006:000B:4
C   006:000D:0
C   006:0010:4
C   006:0011:5
C   006:0013:0
C   006:0016:4
C   006:0017:1
C   006:0017:5
C   006:0019:5
C   006:001A:2
SEGMENT 006 IS 0023 LONG
```

CONSTRUCTION OF THE BAYES SEQUENTIAL DESIGN RULE TO
IDENTIFY THE PARAMETERS OF TWO BINOMIAL POPULATIONS

H1 : THETA = (NU2,NU1) VS H2 : THETA = (NU1,NU2)

Q= 0.70 NU1 = 0.20 NU2 = 0.70 TRUNCATED AT 4 OBSERVATIONS LOSS FACTOR = 50.0

NO. OF OBSERVATIONS = 4 NO. OF STATES = 35

STATE	RISK	E(RISK(X))	E(RISK(Y))	MIN. RISK	STOP	EXP	DECN	FG(STATE)
0004	4.7656				1		2	0.0732
0013	10.3366				1		2	0.0353
0022	25.2360				1		1	0.0312
0031	7.6652				1		1	0.0774
0040	4.4202				1		1	0.2892
0103	12.0000				1		2	0.0245
0112	22.0000				1		1	0.0245
0121	6.8421				1		1	0.0665
0130	4.3208				1		1	0.2525
0202	19.0000				1		1	0.0196
0211	6.1951				1		1	0.0574
0220	4.2448				1		1	0.2206
0301	5.6901				1		1	0.0497
0310	4.1867				1		1	0.1928
0400	4.1424				1		1	0.1686
1003	5.0000				1		2	0.0840
1012	12.0000				1		2	0.0420
1021	22.0000				1		1	0.0420
1030	6.8421				1		1	0.1140
1102	14.0000				1		2	0.0294
1111	19.0000				1		1	0.0336
1120	6.1951				1		1	0.0984
1201	16.3077				1		1	0.0273
1210	5.6901				1		1	0.0852
1300	5.2982				1		1	0.0740
2002	5.3043				1		2	0.0966
2011	14.0000				1		2	0.0504
2020	19.0000				1		1	0.0576
2101	16.3529				1		2	0.0357
2110	16.3077				1		1	0.0468
2200	13.9611				1		1	0.0385
3001	5.6981				1		2	0.1113
3010	16.3529				1		2	0.0612
3100	19.0512				1		2	0.0439
4000	6.2054				1		2	0.1285

NO. OF OBSERVATIONS = 3 NO. OF STATES = 20

STATE	RISK	E(RISK(X))	E(RISK(Y))	MIN. RISK	STOP	EXP	DECN	FG(STATE)
0003	5.5806	6.5806	9.0786	5.5806	1		2	0.1085
0012	19.8421	15.6842	19.7467	15.6842	0	1		0.0665
0021	11.7097	12.7097	18.4344	11.7097	1		1	0.1085
0030	4.1050	5.1050	6.6558	4.1050	1		1	0.3665
0102	23.0000	16.0000	18.0000	16.0000	0	1		0.0490
0111	9.9231	10.9231	16.4032	9.9231	1		2	0.0910
0120	3.8464	4.8464	6.0644	3.8464	1		2	0.3190
0201	8.4545	9.4545	14.4601	8.4545	1		1	0.0770
0210	3.6475	4.6475	5.5974	3.6475	1		1	0.2780
0300	3.4948	4.4948	5.2317	3.4948	1		1	0.2425
1002	6.3333	7.3333	10.3667	6.3333	1		2	0.1260
1011	23.0000	16.0000	18.0000	16.0000	0	1		0.0840
1020	9.9231	10.9231	16.4032	9.9231	1		2	0.1560
1101	26.3333	16.3333	16.8333	16.3333	0	1		0.0630
1110	8.4545	9.4545	14.4601	8.4545	1		1	0.1320
1200	7.2667	8.2667	12.6694	7.2667	1		1	0.1125
2001	7.2857	8.2857	11.8882	7.2857	1		2	0.1470
2010	26.3333	16.3333	16.8333	16.3333	0	1		0.1080
2100	26.2727	16.6727	16.3288	16.3288	0	2		0.0825
3000	8.4783	9.4783	13.6383	8.4783	1		1	0.1725

NO. OF OBSERVATIONS = 2 NO. OF STATES = 10

STATE	RISK	E(RISK(X))	E(RISK(Y))	MIN. RISK	STOP	EXP	DECN	FG(STATE)
0002	10.0000	9.0400	12.8552	9.0400	0	1		0.1750
0011	20.0000	12.8400	13.7764	12.8400	0	1		0.1750
0020	4.8421	5.8421	9.2122	4.8421	1		1	0.4750
0101	17.0000	12.0000	13.2654	12.0000	0	1		0.1400
0110	4.1951	5.1951	7.9667	4.1951	1		1	0.4100
0200	3.6901	4.6901	6.9312	3.6901	1		1	0.3550
1001	12.0000	10.0000	13.1000	10.0000	0	1		0.2100
1010	17.0000	12.0000	13.2654	12.0000	0	1		0.2400
1100	14.3077	11.1007	13.0000	11.1007	0	1		0.1950
2000	14.3529	11.0182	13.4062	11.0182	0	1		0.2550

NO. OF OBSERVATIONS = 1 NO. OF STATES = 4

STATE	RISK	E(RISK(X))	E(RISK(Y))	MIN. RISK	STOP	EXP	DECN	FG(STATE)
0001	21.0000	10.8000	11.3200	10.8000	0	1		0.3500
0010	7.9231	7.0769	9.8869	7.0769	0	1		0.6500
0100	6.4545	6.3175	9.2328	6.3175	0	1		0.5500
1000	24.3333	11.0539	11.1333	11.0539	0	1		0.4500

NO. OF OBSERVATIONS = 0 NO. OF STATES = 1

STATE	RISK	E(RISK(X))	E(RISK(Y))	MIN. RISK	STOP	EXP	DECN	FG(STATE)
0000	15.0000	8.4489	9.1246	8.4489	0	1		1.0000

BAYES RISK = 8.4489

ONESTEPAHEAD

```
C     THIS PROGRAM SIMULATES THE PERFORMANCE OF THE OSA EXPERIMENT RULE
C     FOR THE PROBLEM OF IDENTIFYING ONE OF TWO HYPOTHESES CONCERNING
C     THE PARAMETERS OF TWO BINOMIAL POPULATIONS.  THE ANALYTIC
C     APPROXIMATION FOR THE EXPECTED SAMPLE SIZE USING THE OBSERVED
C     VALUE OF THE P(CS), AND USING PS, IS ALSO COMPUTED.
C     NO = NUMBER OF SIMULATIONS
C     KD = THE SAMPLE OUTPUT (OR SAMPLE SIZE IF THIS IS TOO LARGE) OF
C     EVERY N-TH SIMULATION, WHERE N IS A MULTIPLE OF KD, IS PRINTED
C     PS = P-STAR SIGNIFICANCE LEVEL
C     C = PRIOR PROBABILITY
C     EM1,EM2 = BINOMIAL PARAMETER VALUES

      DIMENSION M(4),G(4,2),KS(1000),A(4),P(4),S(4)
      NO=800
      KD=40
      PS=0.95
      C=.7
      EM1=0.4
      EM2=0.55
   70 DEM1=1.-EM1
      DEM2=1.-EM2
      REM1=EM1/EM2
      RDM1=DEM1/DEM2
      ELOG=ALOG(REM1)
      FLOG=ALOG(RDM1)
      B=ALOG((1.-C)/C)
      NSEED1=-1048477
      NSEED2=-2097149
      NSEED3=-1048295
      NSEED4=-2097051
      NSEED5=-5524519
      WRITE(6,200)

      WRITE(6,250)
      WRITE(6,270)
      WRITE(6,300)C,EM1,EM2,NO,KD
      WRITE(6,350)
      LU=0
      LV=0
      LP=0
      LX=0
      LY=0
      LQ=0
      LR=0
      DO 15 KC=1,NO
      M(1)=0.
      M(2)=0.
      M(3)=0.
      M(4)=0.
      IL=1
      GL=RANDOM(NSEED1)
      IF(GL.LT.C)IS=1
      IF(GL.GE.C)IS=2
   16 CONTINUE
      DO 5 I=1,4
      M(I)=M(I)+1
      A(I)=C*RDM1**(M(1)-M(3))*REM1**(M(2)-M(4))
      G(I,2)=(1.-C)/((1.-C)+A(I))
      G(I,1)=1.-G(I,2)
```

```
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
C  000:00001:5
  START OF SEGMENT 002
FORMAT SEGMENT IS 00D3 LONG
C  002:00000:0
C  002:00000:0
C  002:00001:0
C  002:00001:5
C  002:00003:3
C  002:00005:3
C  002:00007:3
C  002:00009:3
C  002:0000A:4
C  002:0000B:5
C  002:0000D:1
C  002:0000E:3
C  002:00010:0
C  002:00011:3
C  002:00013:5
C  002:00015:4
C  002:00017:4
C  002:00019:4
C  002:0001B:4
C  002:0001D:4
     FIB IS 0006 LONG
C  002:00021:2
C  002:00025:2
C  002:00029:2
C  002:00036:2
C  002:0003A:2
C  002:00035:0
C  002:00038:4
C  002:0003C:2
C  002:0003D:0
C  002:0003D:4
C  002:0003E:2
C  002:0003F:0
C  002:00040:0
C  002:00041:1
C  002:00042:2
C  002:00043:4
C  002:00045:0
C  002:00045:4
C  002:00047:2
C  002:00049:0
C  002:0004B:1
C  002:0004B:1
C  002:0004C:0
C  002:0004E:1
C  002:00055:3
C  002:00059:4
C  002:0005C:3
C  002:0005C:3
C  002:0005C:3
C  002:0005C:3
```

```
      M(I)=M(I)-1
      P(1)=(C/(1.-C))*RDM1**(M(1)-M(3))*REM1**(M(2)-M(4))
      P(2)=1./P(1)
      M(I)=M(I)+1
  105 IF((M(1)-M(3))*FLOG+(M(2)-M(4))*ELOG.GE.B)GO TO 101
      DEN=1.+P(1)
      IF(I.EQ.1)S(I)=DEM2/DEN
      IF(I.EQ.2)S(I)=EM2/DEN
      IF(I.EQ.3)S(I)=DEM1/DEN
      IF(I.EQ.4)S(I)=EM1/DEN
      GO TO 5
  101 DEM=1.+P(2)
      IF(I.EQ.1)S(I)=DEM1/DEM
      IF(I.EQ.2)S(I)=EM1/DEM
      IF(I.EQ.3)S(I)=DEM2/DEM
      IF(I.EQ.4)S(I)=EM2/DEM
    5 M(I)=M(I)-1
      GX=S(1)+S(2)
      GY=S(3)+S(4)
      IF(GX-GY)12,23,32
   32 GR=RANDCM(NSEED3)
      IF(MOD(KC,KD).EQ.0)KS(IL)=0
      IF(GR.GE.EM1.AND.GL.LT.C.OR.GR.GE.EM2.AND.GL.GE.C)IND=1
      IF(GR.LT.EM1.AND.GL.LT.C.OR.GR.LT.EM2.AND.GL.GE.C)IND=2
      GO TO 17
   12 GS=RANDCM(NSEED4)
      IF(MOD(KC,KD).EQ.0)KS(IL)=1
      IF(GS.GE.EM1.AND.GL.GE.C.OR.GS.GE.EM2.AND.GL.LT.C)IND=3
      IF(GS.LT.EM1.AND.GL.GE.C.OR.GS.LT.EM2.AND.GL.LT.C)IND=4
      GO TO 17
   23 GP=RANDCM(NSEED5)
      IF(GP.LT.0.5)GO TO 32
      IF(GP.GE.0.5)GO TO 12
   17 IF(MOD(KC,KD).EQ.0)IL=IL+1
      M(IND)=M(IND)+1
      GACTM=AMAX1(G(IND,1),G(IND,2))
      IF(GACTM.LE.PS)GO TO 16
   56 MSUM=M(1)+M(2)+M(3)+M(4)
      MSUM2=MSUM*MSUM
      IF(MOD(KC,KD).EQ.0.AND.MSUM.GE.132)WRITE(6,900)MSUM
  900 FORMAT(/1X,I3)
      IF(MOD(KC,KD).EQ.0)WRITE(6,100)(KS(II),II=1,MSUM)
      DO 18 J=1,2
      IF(G(IND,J).GE.GACTM)IT=J
   18 CONTINUE
      IF(IT-IS)2,3,2
    2 LS=0
      GO TO 8
    3 LS=1
    8 LU=LU+LS
      MXSUM=M(1)+M(2)
      MYSUM=M(3)+M(4)
      MXSUM2=MXSUM*MXSUM
      MYSUM2=MYSUM*MYSUM
      LV=LV+MSUM
      LP=LP+MSUM2
      LX=LX+MXSUM
      LY=LY+MYSUM
      LQ=LQ+MXSUM2
      LR=LR+MYSUM2
   15 CONTINUE
      ALU=LU
      ALV=LV
      ALP=LP
      ALX=LX
      ALY=LY
      ALQ=LQ
      ALR=LR
      ANO=NO
      PCS=ALU/ANO
      ASN=ALV/ANO
      ASNX=ALX/ANO
      ASNY=ALY/ANO
```

```
C 0021:005C:3
C 0021:005C:3
C 0021:005C:3
C 0021:005C:3
C 0021:005C:3
C 0021:005C:3
C 0021:005C:3
C 0021:005E:4
C 0021:0066:2
C 0021:0068:0
C 0021:006A:1
C 0021:006F:0
C 0021:0070:2
C 0021:0073:4
C 0021:0077:1
C 0021:007A:4
C 0021:007E:1
C 0021:007E:4
C 0021:0080:0
C 0021:0083:2
C 0021:0086:5
C 0021:008A:2
C 0021:008D:5
C 0021:0092:1
C 0021:0093:5
C 0021:0095:5
C 0021:0099:1
C 0021:009A:5
C 0021:009E:1
C 0021:00A3:1
C 0021:00A8:2
C 0021:00AB:5
C 0021:00AA:3
C 0021:00AD:5
C 0021:00B3:0
C 0021:00B8:1
C 0021:00B8:4
C 0021:00BA:2
C 0021:00BC:4
C 0021:00BF:4
C 0021:00C2:5
C 0021:00C5:0
C 0021:00C9:2
C 0021:00CA:2
C 0021:00CD:5
C 0021:00CE:5
C 0021:00D8:2
C 0021:00D8:2
C 0021:00E6:2
C 0021:00E7:0
C 0021:00EB:0
C 0021:00ED:1
C 0021:00EE:3
C 0021:00EF:1
C 0021:00EF:4
C 0021:00F0:2
C 0021:00F1:5
C 0021:00F3:4
C 0021:00F5:5
C 0021:00F7:1
C 0021:00F8:3
C 0021:00FA:0
C 0021:00FB:3
C 0021:00FD:0
C 0021:00FE:3
C 0021:0100:0
C 0021:0101:3
C 0021:0103:4
C 0021:0104:3
C 0021:0105:2
C 0021:0106:1
C 0021:0107:0
C 0021:0107:5
C 0021:0108:4
C 0021:0109:3
C 0021:010A:2
C 0021:010B:4
C 0021:010D:0
C 0021:010E:2
C 0021:010F:4
C 0021:010F:4
C 0021:010F:4
C 0021:010F:4
```

```
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
      SD=SQRT(ALP/ANO-ASN*ASN)
      SDX=SQRT(ALQ/ANO-ASNX*ASNX)
      SDY=SQRT(ALR/ANO-ASNY*ASNY)
      SE=SD/SQRT(ANO)
      SEX=SDX/SQRT(ANO)
      SEY=SDY/SQRT(ANO)
      BS=PCS
      TOP=(2.*PS-1.)*ALOG(PS/(1.-PS))+(2.*C-1.)*B
      TOO=(2.*BS-1.)*ALOG(BS/(1.-BS))+(2.*C-1.)*B
      BOT=0.5*(EM1-EM2)*ALOG(EM1*DEM2/(EM2*DEM1))
      OASN=TOO/BOT
      PASN=TOP/BOT
      WRITE(6,400)PCS,PS
      WRITE(6,420)PASN,OASN
      WRITE(6,450)ASN,ASNX,ASNY
      WRITE(6,500)SD,SDX,SDY
      WRITE(6,550)SE,SEX,SEY
100   FORMAT(/1X,132I1)
200   FORMAT(33X,'MONTE CARLO SIMULATION OF A OSA SEQUENTIAL DESIGN RULE
     1')
250   FORMAT(/28X,'AS APPLIED TO A SIMPLE 2-POINT PRIORS/ BINOMIAL DISTR
     1IBUTIONS MODEL')
270   FORMAT(//35X,'H1 : THETA = (NU1,NU2)    VS    H2 : THETA = (NU2,NU1
     1)')
300   FORMAT(///6X,'G =',F4.1,4X,'NU1 =',F5.2,4X,'NU2 =',F5.2,6X,I5,1X,'
     1SIMULATIONS OF WHICH EVERY',1X,I3,'TH SAMPLE SEQUENCE IS LISTED BE
     1LOW')
350   FORMAT(/5X,'A ZERO DENOTES AN OBSERVATION ON X, AND A ONE DENOTES
     1AN OBSERVATION ON Y')
400   FORMAT(///20X,'OBSERVED PCS =',F11.8,10X,'PCS REQUIRED =',F6.3)
420   FORMAT(//20X,'PREDICTED ASN USING REQUIRED PCS=',F8.3,10X,'PREDICT
     1ED ASN USING OBSERVED PCS=',F8.3)
450   FORMAT(//20X,'OBSERVED ASN=',F8.3,5X,'OBSERVED ASN ON X =',F8.3,5X
     1,'OBSERVED ASN ON Y =',F8.3)
500   FORMAT(//10X,'STANDARD DEVIATION =',F8.5,4X,'STANDARD DEVIATION(X)
     1 =',F8.5,4X,'STANDARD DEVIATION(Y) =',F8.5)
550   FORMAT(//12X,'STANDARD ERROR =',F8.5,8X,'STANDARD ERROR(X) =',F8.5
     1,8X,'STANDARD ERROR(Y) =',F8.5)
      STOP
      END
```

```
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:010F:4
C    002:0112:3
C    002:0115:2
C    002:0118:1
C    002:011A:1
C    002:011C:1
C    002:011E:1
C    002:011F:0
C    002:0124:3
C    002:012A:0
C    002:0130:1
C    002:0131:3
C    002:0132:5
C    002:013B:2
C    002:0143:2
C    002:014D:2
C    002:0157:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0161:2
C    002:0162:1
SEGMENT 002 IS 0162 LONG
```

MONTE CARLO SIMULATION OF A OSA SEQUENTIAL DESIGN RULE
AS APPLIED TO A SIMPLE 2-POINT PRIORS/ BINOMIAL DISTRIBUTIONS MODEL

H1 : THETA = (NU1,NU2) VS H2 : THETA = (NU2,NU1)

G = 0.7 NU1 = 0.40 NU2 = 0.55 800 SIMULATIONS OF WHICH EVERY 40TH SAMPLE SEQUENCE IS LISTED BELOW

A ZERO DENOTES AN OBSERVATION ON X, AND A ONE DENOTES AN OBSERVATION ON Y

```
1111010110101010001011110001100110100000011011001001011101001101010000111110101001101010001010100111000101100110010
1111100011000110101
110011101
205
1001110000110101110011100100110100010101111001101010110001001100010010000110101011010101110011000101010101101010001101011001010101011
0101011111011001100001111111011001110100100000100110010000001101101101110 1
1110101101100100101110000011001010110101001101100 1101
1101010000110100101101011010110 10
1101010011010101010101010101010101000101110001110000101110110101100111001110011110001010110101001000011010001110
1110100
1101010101010101100100010010101010101010011001010010011101011101010101110011011001011101010011000101010010010110001101010111010110 0
1011011001100110000100010100011100110000011111011111110100011000101100 1
1110010101010
10110010000101011
1101110011010100110000011011010110010100000011010101
1001010111010100010100011011100111000100110110000111110011110101010011 01
134
1101110001110101100110101010000101100000010000000101101000110010001000010010101010100111101010000101101000010111101000001101101110111
10
1101000011101100011010011110000111
1011100001111001100110101000100010010111111010000101011101 00
1111110110101C000100101010000101101001 00
1110010101111C000110100100111010101010001110010
187
1110001010001011110100100001010101011010001001100001100111010111110111110101001101001101010001000010011011101110111010101001101100101
0101010110010001111001101010101010101010010101100010111
```

OBSERVED PCS = 0.95750000 PCS REQUIRED = 0.950

PREDICTED ASN USING REQUIRED PCS= 50.837 PREDICTED ASN USING OBSERVED PCS= 55.238

OBSERVED ASN= 51.603 OBSERVED ASN ON X = 25.660 OBSERVED ASN ON Y = 25.943

STANDARD DEVIATION =41.48086 STANDARD DEVIATION(X) =21.28695 STANDARD DEVIATION(Y) =20.42943

```
B6700   F O R T R A N   C O M P I L A T I O N   M A R K   2.9.190   FRIDAY, 11/17/78   04:18 PM
                         O N E S T E P A H E A D
                         = = = = = = = = = = = =
C     THIS PROGRAM SIMULATES THE PERFORMANCE OF THE OSA EXPERIMENT RULE
C     FOR THE PROBLEM OF IDENTIFYING ONE OF TWO HYPOTHESES CONCERNING
C     THE MEANS AND VARIANCES OF TWO NORMAL POPULATIONS.
C     NO = NUMBER OF SIMULATIONS
C     KD : THE SAMPLE SIZE OF EVERY N-TH SIMULATION, WHERE N IS A
C     MULTIPLE OF KD, IS PRINTED
C     PS = P-STAR SIGNIFICANCE LEVEL
C     C = PRIOR PROBABILITY
C     EM1,EM2 = POPULATION MEANS (SYMMETRIC ABOUT 0 WITHOUT LOSS OF
C     GENERALITY)
C     SIG1,SIG2 = POPULATION STANDARD DEVIATIONS
C
      DIMENSION G(2)
      NO=400
      KD=30
      PS=0.95
      C=.5
      EM1=-.25
      EM2=.25
      SEM1=EM1*EM1
      SEM2=EM2*EM2
      SIG1=2.
      SIG2=3.
      VAR1=SIG1*SIG1
      VAR2=SIG2*SIG2
      DIG2=-VAR1*VAR2
      A=VAR2-VAR1
      B=VAR1*EM2-VAR2*EM1
      NSEED1=-1048477
      NSEED6=-524519
      NSEEDX=-2097149
      NSEEDY=-1048295
      WRITE(6,200)

      WRITE(6,250)
      WRITE(6,270)
      WRITE(6,300)C,EM1,EM2,SIG1,SIG2
      WRITE(6,320)NO,KD
      WRITE(6,350)
      LA=0
      LB=0
      LU=0
      LV=0
      LP=0
      LX=0
      LY=0
      LQ=0
      LR=0
      ALK=0.
      ALL=0.
      SIG=1.
      G(1)=C
      G(2)=1.-G(1)
      NMIN=100000
      NMAX=0
      DO 15 KC=1,NO
      GL=RANDCM(NSEED1)
      IF(GL.LT.C)IS=1
      IF(GL.GE.C)IS=2
C
```

```
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
C   000:0000:5
  START OF SEGMENT 002
FORMAT SEGMENT IS 00D9 LONG
C   002:0000:0
C   002:0000:0
C   002:0001:0
C   002:0001:5
C   002:0003:3
C   002:0005:3
C   002:0007:4
C   002:0009:3
C   002:000A:4
C   002:000B:5
C   002:000C:4
C   002:000D:3
C   002:000E:4
C   002:000F:5
C   002:0011:2
C   002:0012:4
C   002:0015:0
C   002:0017:4
C   002:0019:4
C   002:001B:4
C   002:001D:4
        FIB IS 0006 LONG
C   002:0021:2
C   002:0025:2
C   002:0029:2
C   002:0036:2
C   002:003E:2
C   002:0042:2
C   002:0043:0
C   002:0043:4
C   002:0044:2
C   002:0045:0
C   002:0045:4
C   002:0046:2
C   002:0047:0
C   002:0047:4
C   002:0048:2
C   002:0049:0
C   002:0049:4
C   002:004A:2
C   002:004B:3
C   002:004D:1
C   002:004F:3
C   002:0050:1
C   002:0051:0
C   002:0052:4
C   002:0054:2
```

```
C
C
C
C
C
C
      MSUMX=0
      MSUMY=0
      MRAND=0
      SUMX=0.
      SUMY=0.
      SUMSQX=C.
      SUMSQY=C.
   16 D1=2.*VAR1*VAR2*(ALOG(C/(1.-C))+(MSUMX-MSUMY+1)*ALOG(SIG1/SIG2))
      D2=2.*VAR1*VAR2*(ALOG(C/(1.-C))+(MSUMX-MSUMY-1)*ALOG(SIG1/SIG2))
      E=A*(SUMSQX-SUMSQY)
      F=2.*B*(SUMX-SUMY)
      H1=(VAR1*SEM2-VAR2*SEM1)*(MSUMY-MSUMX-1)
      H2=(VAR1*SEM2-VAR2*SEM1)*(MSUMY-MSUMX+1)
      C1=D1+E+F+H1
      Z1=D2+E+F+H2
      C2=-Z1
      IF((B*B-A*C1).LT.0.)GO TO 61
      AL1=(-B-SQRT(B*B-A*C1))/A
      AL2=(-B+SQRT(B*B-A*C1))/A
      XORD1=(AL1-EM2)/SIG2
      XORD2=(AL2-EM2)/SIG2
      XORD3=(AL1-EM1)/SIG1
      XORD4=(AL2-EM1)/SIG1
      IF(VAR1-VAR2)65,90,66
   65 CALL NDTR(XORD1,XP1,Q1)
      CALL NDTR(-XORD2,XP2,Q2)
      CALL NDTR(XORD3,XP3,Q3)
      CALL NDTR(XORD4,XP4,Q4)
      XCF1=XP1+XP2
      XCF2=XP4-XP3
      GO TO 70
   66 CALL NDTR(XORD1,XP1,Q1)
      CALL NDTR(XORD2,XP2,Q2)
      CALL NDTR(XORD3,XP3,Q3)
      CALL NDTR(-XORD4,XP4,Q4)
      XCF1=XP2-XP1
      XCF2=XP3+XP4
      GO TO 70
   61 IF(VAR1-VAR2)67,90,68
   67 XCF1=1.
      XCF2=0.
      GO TO 70
   68 XCF1=0.
      XCF2=1.
   70 IF((B*B-A*C2).LT.0.)GO TO 62
      BL1=(-B-SQRT(B*B-A*C2))/A
      BL2=(-B+SQRT(B*B-A*C2))/A
      YORD1=(BL1-EM1)/SIG1
      YORD2=(BL2-EM1)/SIG1
      YORD3=(BL1-EM2)/SIG2
      YORD4=(BL2-EM2)/SIG2
      IF(VAR1-VAR2)63,90,64
   63 CALL NDTR(YORD1,YP1,R1)
      CALL NDTR(YORD2,YP2,R2)
      CALL NDTR(YORD3,YP3,R3)
      CALL NDTR(-YORD4,YP4,R4)
      YCF1=YP2-YP1
      YCF2=YP3+YP4
      GO TO 80
   64 CALL NDTR(YORD1,YP1,R1)
      CALL NDTR(-YORD2,YP2,R2)
      CALL NDTR(YORD3,YP3,R3)
      CALL NDTR(YORD4,YP4,R4)
      YCF1=YP1+YP2
      YCF2=YP4-YP3
      GO TO 80
   62 IF(VAR1-VAR2)69,90,71
   69 YCF2=1.
      YCF1=0.
      GO TO 80
   71 YCF2=0.
      YCF1=1.
C
C
C
C
```

```
C    002:0056:3
C    002:0056:3
C    002:0056:3
C    002:0056:3
C    002:0056:3
C    002:0056:3
C    002:0056:3
C    002:0057:1
C    002:0057:5
C    002:0058:3
C    002:0059:1
C    002:0059:5
C    002:005A:3
C    002:005B:1
C    002:0062:0
C    002:0068:5
C    002:006A:4
C    002:006D:0
C    002:0070:4
C    002:0074:2
C    002:0076:4
C    002:0079:0
C    002:0079:4
C    002:007C:1
C    002:0080:1
C    002:0084:1
C    002:0086:0
C    002:0087:5
C    002:0089:4
C    002:008B:3
C    002:008E:5
C    002:0091:0
C    002:0093:1
C    002:0095:2
C    002:0097:3
C    002:0098:5
C    002:009A:1
C    002:009A:4
C    002:009C:5
C    002:009F:0
C    002:00A1:1
C    002:00A3:2
C    002:00A4:4
C    002:00A6:0
C    002:00A6:3
C    002:00A9:5
C    002:00AA:3
C    002:00AB:1
C    002:00AB:4
C    002:00AC:2
C    002:00AD:0
C    002:00AF:3
C    002:00B3:3
C    002:00B7:3
C    002:00B9:2
C    002:00BB:1
C    002:00BD:0
C    002:00BE:5
C    002:00C2:1
C    002:00C4:2
C    002:00C6:3
C    002:00C8:4
C    002:00CA:5
C    002:00CC:1
C    002:00CD:3
C    002:00CE:0
C    002:00D0:1
C    002:00D2:2
C    002:00D4:3
C    002:00D6:4
C    002:00D8:0
C    002:00D9:2
C    002:00D9:5
C    002:00DD:1
C    002:00DD:5
C    002:00DE:3
C    002:00DF:0
C    002:00DF:4
C    002:00E0:2
C    002:00E0:2
C    002:00E0:2
C    002:00E0:2
```

```
C
C
C
C
C
C
C
C
   80 IF((YCF1-XCF1)*G(1)+(YCF2-XCF2)*G(2))32,23,12
   32 IF(IS.EQ.1)CALL GAUSS(NSEEDX,SIG2,EM2,X)
      IF(IS.EQ.2)CALL GAUSS(NSEEDX,SIG1,EM1,X)
      SUMX=SUMX+X
      SUMSQX=SUMSQX+X*X
      MSUMX=MSUMX+1
      GO TO 17
   12 IF(IS.EQ.1)CALL GAUSS(NSEEDY,SIG1,EM1,Y)
      IF(IS.EQ.2)CALL GAUSS(NSEEDY,SIG2,EM2,Y)
      SUMY=SUMY+Y
      SUMSQY=SUMSQY+Y*Y
      MSUMY=MSUMY+1
      GO TO 17
   23 GP=RANDCM(NSEED6)
      MRAND=MRAND+1
      IF(GP.LT.0.5)GO TO 32
      IF(GP.GE.0.5)GO TO 12
   17 G(1)=1./(1.+((1.-C)/C)*SIG1**(MSUMY-MSUMX)*SIG2**(MSUMX-MSUMY)*EXP
     1((A*(SUMSQX-SUMSQY)-2.*B*(SUMY-SUMX)+(MSUMY-MSUMX)*(VAR1*SEM2-VAR2
     2*SEM1))/(2.*DIG2)))
      G(2)=1.-G(1)
      GACTM=AMAX1(G(1),G(2))
      TM=TIME(2)
      IF(TM.GE.6500)GO TO 56
      IF(GACTM.LE.PS)GO TO 16
   56 MSUM=MSUMX+MSUMY
      MRAND2=MRAND*MRAND
      NMAX=MAXO(NMAX,MSUM)
      NMIN=MINO(NMIN,MSUM)
      IF(MOD(KC,KD).EQ.0)WRITE(6,100)MSUM,MSUMX,MSUMY
      IF(G(1).GE.PS)IT=1
      IF(G(2).GE.PS)IT=2
      IF(IT-IS)2,3,2
    2 LS=0
      GO TO 8
    3 LS=1
    8 LU=LU+LS
      RSUM=MSUM
      SLOG=ALOG(RSUM)
      SQLOG=SLOG*SLOG
      MSUM2=MSUM*MSUM
      MXSUM=MSUMX
      MYSUM=MSUMY
      MXSUM2=MXSUM*MXSUM
      MYSUM2=MYSUM*MYSUM
      LA=LA+MRAND
      LB=LB+MRAND2
      LV=LV+MSUM
      LP=LP+MSUM2
      LX=LX+MXSUM
      LY=LY+MYSUM
      LQ=LQ+MXSUM2
      LR=LR+MYSUM2
      ALK=ALK+SLOG
      ALL=ALL+SQLOG
      TN=TIME(2)
      IF(TN.GE.6500)NO=KC
   15 CONTINUE
      ALA=LA
      ALB=LB
      ALU=LU
      ALV=LV
      ALP=LP
      ALX=LX
      ALY=LY
      ALQ=LQ
      ALR=LR
      ANO=NO
      PCS=ALU/ANO
      ASN=ALV/ANO
      ANR=ALA/ANO
      ALSN=ALK/ANO
      ASNX=ALX/ANO
      [illegible]
      [illegible]
```

```
C     002:00E0:2
C     002:00E0:2
C     002:00E0:2
C     002:00E0:2
C     002:00E0:2
C     002:00E0:2
C     002:00E0:2
C     002:00E0:2
C     002:00E0:2
C     002:00E6:0
C     002:00E9:5
C     002:00ED:5
C     002:00EF:1
C     002:00F0:5
C     002:00F2:1
C     002:00F2:4
C     002:00F6:3
C     002:00FA:3
C     002:00FB:5
C     002:00FD:3
C     002:00FE:5
C     002:00FF:2
C     002:0101:0
C     002:0102:2
C     002:0104:4
C     002:0107:4
C     002:010E:3
C     002:0114:0
C     002:0117:1
C     002:0118:5
C     002:011C:0
C     002:011D:3
C     002:011E:4
C     002:0120:0
C     002:0121:3
C     002:0122:5
C     002:0125:5
C     002:0128:5
C     002:0134:2
C     002:0136:3
C     002:0138:5
C     002:013A:1
C     002:013A:5
C     002:013B:2
C     002:013C:0
C     002:013D:3
C     002:013E:2
C     002:013F:5
C     002:0140:4
C     002:0142:0
C     002:0142:5
C     002:0143:4
C     002:0145:0
C     002:0146:2
C     002:0147:5
C     002:0149:2
C     002:014A:5
C     002:014C:2
C     002:014D:5
C     002:014F:2
C     002:0150:5
C     002:0152:2
C     002:0153:4
C     002:0155:0
C     002:0156:3
C     002:0158:3
C     002:015A:4
C     002:015B:3
C     002:015C:2
C     002:015D:1
C     002:015E:0
C     002:015E:5
C     002:015F:4
C     002:0160:3
C     002:0161:2
C     002:0162:1
C     002:0163:0
C     002:0164:2
C     002:0165:4
C     002:0167:0
C     002:0168:2
C     [illegible]
```

```
C
C
C
C
C
C
C
C
      SD=SQRT(ALP/ANO-ASN*ASN)
      SDL=SQRT(ALL/ANO-ALSN*ALSN)
      SDX=SQRT(ALQ/ANO-ASNX*ASNX)
      SDY=SQRT(ALR/ANO-ASNY*ASNY)
      SDR=SQRT(ALB/ANO-ANR*ANR)
      SE=SD/SQRT(ANO)
      SEL=SDL/SQRT(ANO)
      SEX=SDX/SQRT(ANO)
      SEY=SDY/SQRT(ANO)
      SER=SDR/SQRT(ANO)
      WRITE(6,390)NO
      WRITE(6,400)PCS,PS
      WRITE(6,450)ASN,ASNX,ASNY
      WRITE(6,460)ANR,SDR,SER
      WRITE(6,470)NMIN,NMAX
      WRITE(6,500)SD,SDX,SDY
      WRITE(6,550)SE,SEX,SEY
      WRITE(6,600)ALSN,SDL,SEL
      GO TO 81
   90 WRITE(6,1000)
 1000 FORMAT(///40X,33('*'),/20X,'*ERROR : VAR1 AND VAR2 ARE EQUAL*',/20
     1X,33('*'))
  100 FORMAT(/24X,I4,17X,I4,20X,I4)
  200 FORMAT(30X,'MONTE CARLO SIMULATION OF AN OSA SEQUENTIAL DESIGN RUL
     1E')
  250 FORMAT(/28X,'AS APPLIED TO A SIMPLE 2-POINT PRIORS/NORMAL DISTRIBU
     1TIONS MODEL')
  270 FORMAT(//1X,'H1 : THETA = (NU2,VAR2,NU1,VAR1)  VS  H2 : THETA = (N
     1U1,VAR1,NU2,VAR2)  WHERE NUX,VARX ARE PARAMETERS OF THE 1ST POPULA
     2TION')
  300 FORMAT(///8X,'G =',F4.1,4X,'NU1 =',F5.2,4X,'NU2 =',F5.2,4X,'SIGMA1
     1 =',F4.1,4X,'SIGMA2 =',F4.1)
  320 FORMAT(//10X,I5,2X,'SIMULATIONS OF WHICH EVERY',1X,I3,'TH SAMPLE S
     1IZE IS LISTED BELOW')
  350 FORMAT(///20X,'SAMPLE SIZE',10X,'SAMPLE SIZE(X)',10X,'SAMPLE SIZE(
     1Y)',/19X,13('*'),8X,16('*'),8X,16('*'))
  390 FORMAT(///30X,'SIMULATIONS TRUNCATED TO',1X,I3,1X,'IF PROCESSTIME
     1IS TOO LARGE')
  400 FORMAT(///20X,'OBSERVED PCS =',F11.8,10X,'PCS REQUIRED =',F6.3)
  450 FORMAT(//20X,'OBSERVED ASN=',F8.3,5X,'OBSERVED ASN ON X =',F8.3,5X
     1,'OBSERVED ASN ON Y =',F8.3)
  460 FORMAT(///15X,'OBSERVED AN OF RANDOMISATIONS = ',F8.3,5X,'STANDARD
     1 DEVIATION = ',F8.5,4X,'STANDARD ERROR = ',F8.5)
  470 FORMAT(///20X,'RANGE OF ASN = (',I3,',',I3,')')
  500 FORMAT(//10X,'STANDARD DEVIATION =',F8.5,4X,'STANDARD DEVIATION(X)
     1 =',F8.5,4X,'STANDARD DEVIATION(Y) =',F8.5)
  550 FORMAT(//12X,'STANDARD ERROR =',F8.5,8X,'STANDARD ERROR(X) =',F8.5
     1,8X,'STANDARD ERROR(Y) =',F8.5)
  600 FORMAT(//20X,'OBSERVED AVERAGE LOG SAMPLE SIZE = ',F6.3,5X,'STANDA
     1RD DEVIATION = ',F6.3,10X,'STANDARD ERROR = ',F7.4)
   81 STOP
      END
```

```
C    002:016B:0
C    002:016B:0
C    002:016B:0
C    002:016B:0
C    002:016B:0
C    002:016B:0
C    002:016B:0
C    002:016B:0
C    002:016B:0
C    002:016D:5
C    002:0170:4
C    002:0173:3
C    002:0176:2
C    002:0179:1
C    002:017B:1
C    002:017D:1
C    002:017F:1
C    002:0181:1
C    002:0183:1
C    002:018A:2
C    002:0192:2
C    002:019C:2
C    002:01A6:2
C    002:01AE:2
C    002:01B8:2
C    002:01C2:2
C    002:01CC:2
C    002:01CC:5
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D0:2
C    002:01D1:1
SEGMENT 002 IS 0200 LONG
```

```
      SUBROUTINE NDTR(X,P,D)
C
C     THIS SUBROUTINE CALCULATES THE STANDARD NORMAL DENSITY VALUE D AND
C
C     THE CUMULATIVE DISTRIBUTION VALUE P FOR THE VALUE X.
C
      AX=ABS(X)
      T=1./(1.+.2316419*AX)
      D=0.39894228*EXP(-X*X/2.)
      P=1.-D*T*((((1.330274429*T-1.821255978)*T+1.781477937)*T-0.3565637
     182)*T+0.31938153)
      IF(X)1,2,2
    1 P=1.-P
    2 RETURN
      END
```

```
 START OF SEGMENT 00A
C   00A:0000:0
C   00A:0000:0
C   00A:0000:0
C   00A:0000:0
C   00A:0000:0
C   00A:0000:0
C   00A:0000:0
C   00A:0001:1
C   00A:0004:2
C   00A:0008:5
C   00A:000F:4
C   00A:0014:0
C   00A:0015:1
C   00A:0016:2
C   00A:0016:5
SEGMENT 00A IS 0020 LONG
```

```
      SUBROUTINE GAUSS (NSEEDG,S,AM,V)
C     THIS SUBROUTINE GENERATES NORMAL RANDOM VARIABLES GIVEN THE MEAN
C     AND VARIANCE
C     AM = MEAN OF NORMAL DISTRIBUTION
C     S = STANDARD DEVIATION OF NORMAL DISTRIBUTION
C     V = OUTPUT NORMAL RANDOM VARIABLE
      A=0.0
      DO 8 I=1,12
      Y=RANDOM(NSEEDG)
    8 A=A+Y
      V=(A-6.0)*S+AM
      RETURN
      END
```

```
 START OF SEGMENT 00B
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:0
C   00B:00001:4
C   00B:00002:0
C   00B:00003:4
C   00B:00007:1
C   00B:00009:3
C   00B:0000A:0
SEGMENT 00B IS 0014 LONG
```

MONTE CARLO SIMULATION OF AN OSA SEQUENTIAL DESIGN RULE
AS APPLIED TO A SIMPLE 2-POINT PRIORS/NORMAL DISTRIBUTIONS MODEL

H1 : THETA = (NU2,VAR2,NU1,VAR1) VS H2 : THETA = (NU1,VAR1,NU2,VAR2) WHERE NUX,VARX ARE PARAMETERS OF THE 1ST POPULATION

G = 0.5 NU1 =-0.25 NU2 = 0.25 SIGMA1 = 2.0 SIGMA2 = 3.0

400 SIMULATIONS OF WHICH EVERY 30TH SAMPLE SIZE IS LISTED BELOW

SAMPLE SIZE	SAMPLE SIZE(X)	SAMPLE SIZE(Y)
41	14	27
11	11	0
16	3	13
18	15	3
23	6	17
17	13	4
10	0	10
32	29	3
25	25	0
14	0	14
2	1	1
6	3	3
22	15	7

SIMULATIONS TRUNCATED TO 400 IF PROCESSTIME IS TOO LARGE

OBSERVED PCS = 0.96250000 PCS REQUIRED = 0.950

OBSERVED ASN= 16.785 OBSERVED ASN ON X = 8.678 OBSERVED ASN ON Y = 8.108

OBSERVED AN OF RANDOMISATIONS = 0.003 STANDARD DEVIATION = 0.04994 STANDARD ERROR = 0.00250

RANGE OF ASN = (1, 51)

STANDARD DEVIATION = 9.33937 STANDARD DEVIATION(X) = 8.00865 STANDARD DEVIATION(Y) = 8.13517

STANDARD ERROR = 0.46697 STANDARD ERROR(X) = 0.40043 STANDARD ERROR(Y) = 0.40676

OBSERVED AVERAGE LOG SAMPLE SIZE = 2.641 STANDARD DEVIATION = 0.671 STANDARD ERROR = 0.0336